Information Sciences Letters
An International Journal

http://dx.doi.org/10.18576/isl/110317

Mouth Image Based Person Authentication Using DWLSTM and GRU

Showkat A. Dar[*], *S. Palanivel*, *M. Kalaiselvi Geetha* and *M. Balasubramanian*

Department of Computer Science and Engineering, Annamalai University, Annamalainagar, 608002 Chennai, India

Received:1 Feb. 2022, Revised: 22 Mar. 2022, Accepted: 24 Mar. 2022.
Published online:1 May 2022.

Abstract: Recently several classification methods were introduced to solve mouth based biometric authentication systems. The results of previous investigations into mouth prints are insufficient and produce lesser authentication results. This is mainly due to the difficulties that accompany any analysis of the mouths: mouths are very flexible and pliable, and successive mouth print impressions even those obtained from the same person may significantly differ from one other. The existing machine learning methods, may not achieve higher performance and only few methods are available using deep learning for mouth biometric authentication. The use of deep learning based mouth biometrics authentication gives higher results than usual machine learning methods. The proposed mouth based biometric authentication (MBBA) system is rigorously examined with real world data and challenges with the purpose that could be expected on mouth-based solution deployed on a mobile device. The proposed system has three major steps such as (1) database collection, (2) creating model for authentication, (3) performance evaluation. The database is collected from Annamalai University deep learning laboratory which consists of 5000 video frames belongs to 10 persons. The person authentication model is created using divergence weight long short term memory (DWLSTM) and gated recurrent unit (GRU) to capture the temporal relationship in mouth images of a person. The existing and proposed methods are implemented via the Anaconda with Jupyter notebook. Finally the results of the proposed model are compared against existing methods such as support vector machine (SVM), and Probabilistic Neural Network (PNN) with respect to metrics like precision, recall, F1-score, and accuracy of mouth.

Keywords: Biometric authentication, deep learning, mouth authentication, divergence weight long short term memory (DWLSTM), gated recurrent unit (GRU).

1 Introduction

Wireless technologies have become increasingly popular in burgeoning military affairs and so is security, where the identity of an individual on the other side of the network has become challenging to determine [1]. Thus, authenticity of an individual to access the private resources is the foremost concern and strong authentication between two parties (end- to-end) needs to be implemented. With mobile devices constantly taking a bigger part in everyday life, the ease of accessing a bank account, paying for any services or even checking medical journals independently of current time and place is getting more and more feasible. Having in mind that these kinds of services require access to the personal information or user, the logical major requirement is high security and strong user authentication methods [1, 2, 3]. Secure authentication before gaining access to personal devices is essential. Biometric authentication is the process of verifying the claimed identification of a person based on an innate human characteristic or trait [4,5]. Biometric authentication is the process of verifying the claimed identification of a person based on an innate human characteristic or trait.

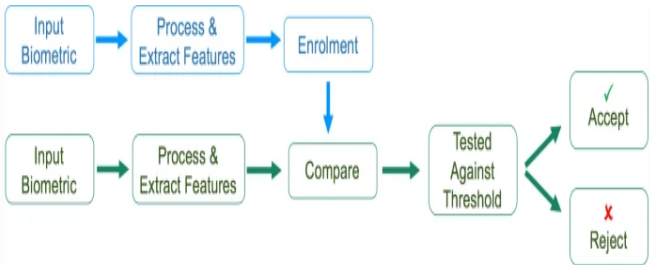


Fig. 1: Biometric authentication overview, involving an enrolment stage shown in blue, and authentication stage shown in green.

Fig. 1 gives an overview of the two stages of biometric authentication which first includes an enrolment

[*]Corresponding author-mail: showkatme2009@gmail.com

stage; users can then authenticate themselves against the enrolment data. However, identification differs from authentication in that it is the ability to identify an individual from a predefined group of users. Biometric traits can be physiological or behavioural. Physiological biometrics such as face or fingerprint have already been successfully rolled out in many state-of-the-art devices, both of these examples have been spoofed in high profile media cases. Behavioural biometrics captures a pattern or behaviour such as signature or voice verification. Behavioural biometrics can be more difficult to spoof; however, they can also be more difficult to model and authenticate robustly. Within biometric authentication, liveness detection refers to being able to detect if a human is live and present during the authentication process. If liveness is successfully incorporated within a biometric system, it could prevent face recognition systems from being spoofed using photographs or artificial fingerprints being successful. Liveness detection is naturally easier to build into a behavioural biometric system as the behaviour requested can be altered.

Mouth based biometric authentication (MBBA) is the process of authenticating a person based on their visual mouth movements while speaking [6, 7]. MBBA has great potential for mobile devices; it is a behavioural biometric in which liveness could be easily incorporated by randomizing the requested spoken content, and it can be captured using a device's front-facing camera. When speaking, people's mouths are involved in motions. Studies show that such motions present unique mouth movement patterns for different individuals. This triggers present research work that is used to extract behavioural patterns of mouth image for user authentication on mobile devices such as smartphones and smart pads. Recently, mouths recognition [8, 9, 10] has been proposed as a new relevant emerging kind of biometrics, which is derived from criminal and forensic real-life applications.

Mouth information has also been used in identification. In [11] mouth features are generated using information about the mouth area, height and width of the mouth contours, oral cavity pixels and visible teeth. Their best recorded result for identification was reported as 94.7% accuracy. The main objective of this work is to authenticate a person from videos by using his/her facial component via mouths. This work is to prove the benefit of mouths as a biometric modality by using both divergence weight long short term memory (DWLSTM) and gated recurrent unit (GRU). Finally, to strengthen the reliability of the authentication results, hybrid classifier is used for user authentication by examining the mouth patterns. The proposed authentication framework is suitable for any camera based devices such as smartphones, tablets, and laptops. Extensive experiments demonstrate that the MBBA system is reliable and efficient for user authentication in real environments.

Wright and Stewart [6] developed to push the field forward through the application of deep learning. A deep artificial neural network (ANN) using spatiotemporal convolutional and bidirectional gated recurrent unit layers is trained end-to-end. For the first time one-shot-learning is applied to mouth-based biometric authentication by implementing a siamese network architecture, meaning the model only needs a single prior example in order to authenticate new users. This approach sets a new modern performance for mouth-based biometric authentication on the XM2VTS dataset and Lausanne protocol with an equal error rate of 0.93% on the evaluation set and a false acceptance rate (FAR) of 1.07% at a 1.00% false rejection rate (FRR).

Wright and Stewart [7] proposed a mouth-based authentication system that performs beyond a closed-set protocol, benchmarking a new open-set protocol with equal error rates of 1.65% on the XM2VTS dataset. New datasets, qFace and FAVMOUTHS, were collected for the work, which push the field forward by enabling systematic testing of the content and quantities of data needed for mouth-based biometric authentication. The FAVMOUTHS dataset was designed to mimic some of the hardest challenges that could be expected in a deployment scenario and include varied spoken content, miming and a wide range of challenging lighting conditions. The datasets captured for this work are available to other university research groups on request.

Wrobel et al [12] proposed a mouth-based biometric recognition approach with the probabilistic neural network (PNN). In the first step, mouth area is restricted to a region of interest (ROI) and in the second step; features extracted from ROI are specifically modelled by dedicated image processing algorithms. Extracted mouth features are then an input data of PNN. All experiments were confirmed in the ten-fold cross validation fashion on three diverse datasets, Multi-PIE Face Dataset, PUT database and own faces dataset. Announced results were verified in the comparative studies and confirm the efficiency of the proposed mouth based biometrics learned by particle swarm optimization (PSO) technique. Results achieved by PNN were improved by the PSO technique.

Cruz et al [13] proposed a mouth biometric system that focuses on the uniqueness of the parameters of the mouths as a useful feature to distinguish similar-looking people. Data gathering includes five identical twins, ten similar faces, and ten dissimilar faces of still face-front images of subjects with neutral expressions were used to examine the efficiency and performance of the system. Different lighting conditions measured in flux under various distances have been characterized. The Viola-Jones algorithm is proposed for face detection and the active appearance model (AAM) for mouth extraction.

Lu et al [14] proposed a mouth reading-based user authentication system, MouthPass, which extracts unique behavioural characteristics of users' speaking mouths leveraging built-in audio devices on smartphones for user authentication. First investigate Doppler profiles of acoustic

signals caused by users' speaking mouths, and find that there are unique mouth movement patterns for different individuals. To characterize the mouth movements, a deep learning-based method is proposed to extract efficient features from Doppler profiles, support vector machine (SVM), and support vector domain description is introduced to construct binary classifiers and spoofer detectors for user identification and spoofer detection, respectively. Afterwards, a binary tree-based authentication approach is proposed to accurately identify each individual leveraging these binary classifiers and spoofer detectors with respect to registered users. Through extensive experiments, MouthPass can achieve 90.21% accuracy in user identification and 93.1% accuracy in spoofer detection.

Wrobel et al [15] proposed a new method of personal identification that analyzes mouth prints. In spite of its important role in forensic and biometric applications, the results of previous investigations into mouth prints are scanty. This pattern contains only such furrows that occur on the greatest number of mouth prints obtained from the same person, where these furrows' locations and inclinations remain similar across the mouth prints obtained. It should be noted that in approach, instead of mouth photos mouth prints are employed which can be obtained at a crime scene. It is worth noticing also that a new method is introduced for personal identification were, instead of popular machine learning methods, the furrow-analysis of mouth prints. According to the authors' convictions, based on reports in the literature, the proposed approach describes for the first time a strategy as to how mouth print structures could be analyzed in biometric applications.

Porwik et al [16] proposed a classification method based on an ensemble of binary classifiers. This strategy consists of two phases: (1) the competence of the base heterogeneous classifiers in a pool is determined, and (2) an ensemble is formed by combining those base classifiers with the greatest competences for the given input data. Results showed that the competence of the base classifiers can be successfully calculated even if the number of their learning examples was limited. Such a situation is particularly observed with biometric data. New biometric data structure is introduced in which the Sim coefficients, along with an efficient data processing technique involving a pool of competent classifiers chosen by dynamic selection. A public dataset retrieved from a biometric repository was used to test the quality of the proposed approach. All the results clearly showed the remarkable performance of ensembles generated by proposed dynamic base-classifier selection, an approach which outperformed all other methods. The proposed ensemble-based strategy demonstrates higher recognition accuracy.

Wright et al [17] investigated using mouth movements as a behavioural biometric for person authentication. The system was trained, evaluated and tested using the XM2VTS dataset, following the Lausanne Protocol configuration II. Features were selected from the discrete cosine transform (DCT) coefficients of the grayscale mouth image. DCT coefficients are selected, the selection process, and static and dynamic feature combinations. Using a Gaussian mixture model (GMM)-universal background model framework an equal error rate of 2.20% was achieved during evaluation and on an unseen test set a false acceptance rate of 1.7% and false rejection rate of 3.0% was achieved. This compares favourably with face authentication results on the same dataset whilst not being susceptible to spoofing attacks.

In the last decade researchers have focused on the use of hybrid intelligent systems for classification. For such systems, various computational-intelligence techniques have been proposed to solve realistically complex problems as seen in the fields of medical diagnosis and biomedical technology, image analysis, biometrics, banking, data analytics and many others. Hybridization techniques have mostly been inspired by the human behavioural system because it integrates information from different parts of the body before finally coordinating the activity of all of the body's parts. In the technical sciences, this approach can be analogously treated as a hybrid composition of many diverse computational units which together help to form an ultimate decision. The major aim of the work is to design a hybrid classification system for biometric authentication of users based on the mouth.

2 Materials and Methods

In this paper, Mouth Based Biometric Authentication (MBBA) is introduced for user authentication and it can be deployed on a mobile device. The proposed system consists three major steps such as (1) database collection (2) creating model for authentication (3) performance evaluation. For the first step, the model created for person authentication is using DWLSTM and GRU as shown in Fig. 2.

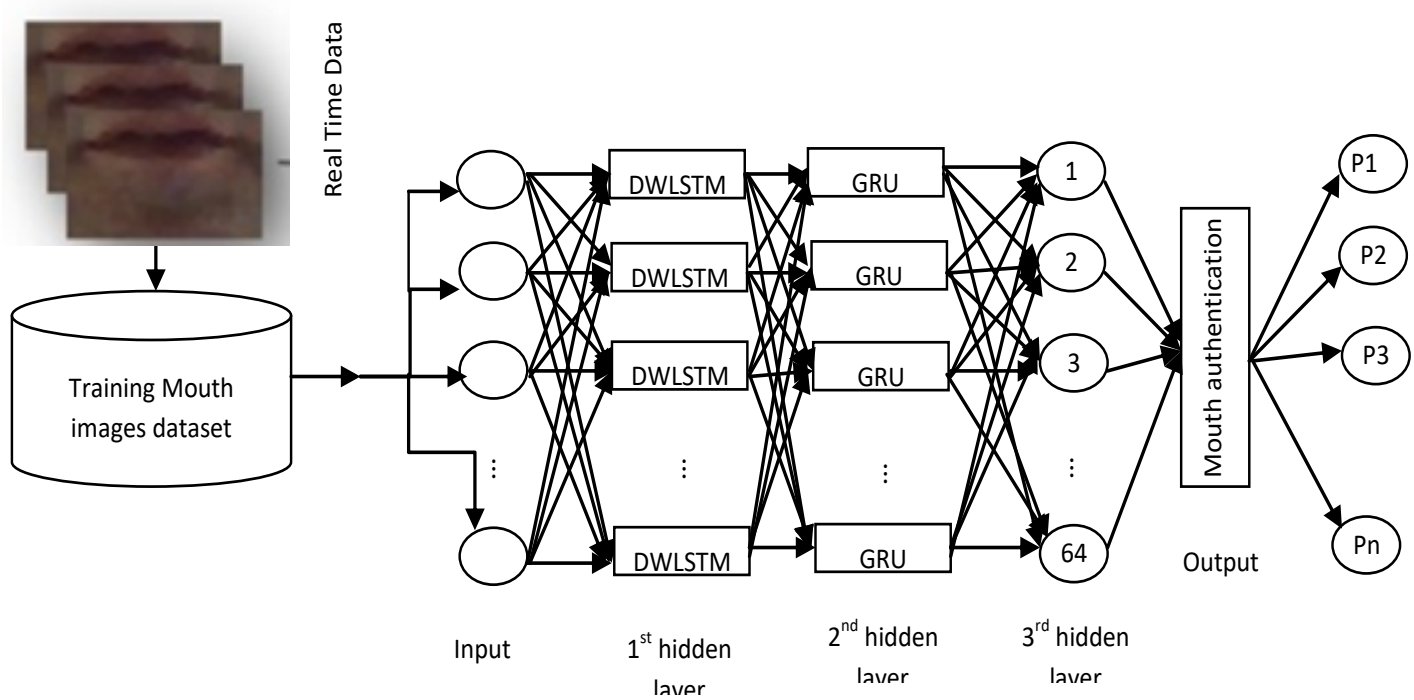


Fig. 2: Architecture of mouth-based biometric authentication (MBBA) system.

In this work, 5000 live mouth video frames in real time for 10 persons are collected with image size of 81x48.Each Person's live mouth of 5000 consecutive video frames extracted.

2.1. Mouth based biometric authentication (MBBA)

At first, all the images of the dataset are used as the input of the DWLSTM layer. DWLSTM is the first hidden layer. Each DWLSTM neuron collects the data and along the path, a weighted value is generated. Image is then passed from the DWLSTM layer to the GRU layer which is the second hidden layer. Again, a weighted value is generated along the path from the DWLSTM layer to the GRU layer. Similarly, data is then passed to the dense layer which is the third hidden layer. A weighted value is generated from GRU to the dense layer. The dense layer is a normal neural network layer that we have used to produce the output. From the third hidden layer, the image is then passed to the output neuron and weight is generated correspondingly. The output is then compared with the original value to find out the error function. The weighted values are then updated according to the difference of the actual value and predicted value until it reaches the minimum point of the cost function and weights are then saved for future predictions and the system's performance is measured. The entire system is trained for 50 epochs.

2.1.1 Divergence weight long short term memory (DWLSTM) classifier

In this work, divergence weight long short term memory (DWLSTM) classifier is introduced to solve the limitations of long-term dependencies in general classifier, a linear self loop memory cell that showed data values through the mouth biometric based authentication [18, 19]. The memory cell is moderated through the amount of mouth biometric flow in and out from the cell. Instead of a simple recurrent neural network (RNN) unit, a DWLSTM unit has a memory cell that has state $c_t \in R^K$ at time t. Through the memory cell, the mouth biometric image is flowed and controlled by three gates such as an input gate secondly a forget gate and finally an output gate. The input gate $x_t \in R^K$ controls the flow of mouth biometric images into the cell. Forgetting memory cell is controlled by the $f_t \in R^K$ forget gate and the output gate $o_t \in R^K$ modifies the output flow from the memory cell. The element wise sigmoid function of a vector by σ and the element wise product of two vectors by the three gated are all the sigmoid units that sets each and every element gate value from 0 and 1 [20,21].

Each memory cell in the DWLSTM neuron stored other mouth biometric image that maintains cell state of its own. Whereas, the neurons in normal RNNs merely consider their previous hidden state and the present output input state to a new hidden state. The DWLSTM neuron considers its old cell state and the outputs of its new cell state. The DWLSTM consists of cell state, input gate, hidden state, Forget gate, and output gate.

Forget gate: The forget gate decides when specific portions of the cell state are to be replaced with more recent information. Its outputs values, close to 1 of the cell state should be retained, and zero for values that should be neglected.

Input gate: Based on the input output o_{t-1}, input x_t , and previous cell state c_{t-1}, the network learns the conditions under any information that should be stored or updated in the cell state.

Hidden state (h_t): It is calculated by multiplying output gate vector by cell state vector

Cell state: 1D vector of fixed shape with random value initialization. It contains the information that was present in the memory after the previous time step. DWLSTM architecture consists of five main parts as shown in Fig. 3.

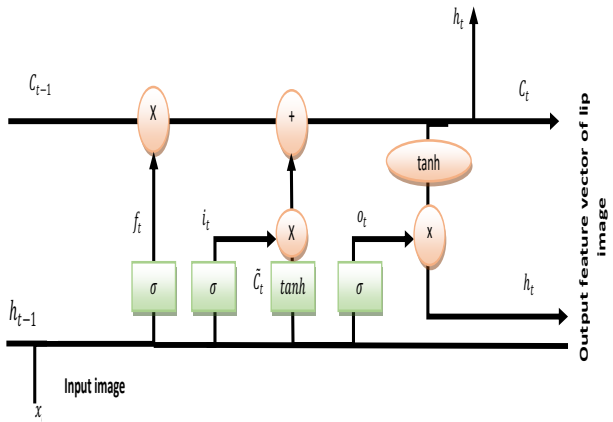


Fig. 3: A single cell DWLSTM architecture.

These three layers play an important role in activation function. The activation function decides whether the neuron should be activated or deactivated. The activation function does the non-linear transformation to the input making it capable to perform more complex tasks. The values of these vectors are calculated by the Equations (1)–(5)[22],

$$i_t = \sigma\left(W^{(i)}x_t + U^{(i)}h_{t-1} + b^{(i)}\right) \quad (1)$$

$$f_t = \sigma\left(W^{(f)}x_t + U^{(f)}h_{t-1} + b^{(f)}\right) \quad (2)$$

$$o_t = \sigma\left(W^{(o)}x_t + U^{(o)}h_{t-1} + b^{(o)}\right) \quad (3)$$

$$c_t = i_t \odot u_t \odot f_t \odot c_{t-1} \quad (4)$$

$$h_t = o_t \odot \tanh(c_t) \quad (5)$$

In the above equations, x_t are input vectors, c_{t-1} signify previous cell states while h_{t-1} are hidden states, W represents input-to-hidden, U stands for hidden-to-hidden weight matrices, σ signifies sigmoid function while ⊙ the operator for element multiplications. Weighing operations are crucial in biometric identifications as images picture have varied significance in relation to target notions. Assuming a given feature value has information for target classes in computing weights for image authentications, differences between prior and posterior class distributions are used to determine the amount of information contained in particular feature values. The range of a feature values are estimated using KL (Kullback-Leibler) depicted in equation (6).

$$KL\left(C|f_{ij}\right) = \sum_{C} P\left(c|f_{ij}\right) \log\left(\frac{P\left(c|f_{ij}\right)}{P(c)}\right) (6)$$

Where f_{ij} implies i^{th}feature's jth value in training biometric image datasets. The feature's weight $fw_{avg}(i)$, defined as weighted averages of feature i's KL values and depicted by equation (7),

$$fw_{avg}(i) = \sum_{j|i} P\left(f_{ij}\right).KL\left(C|f_{ij}\right) \quad (7)$$

In this equation (7), $P(f_{ij})$ means the probability that the feature i has the value of f_{ij}. Above weight $fw_{avg}(i)$ is biased towards feature with many values. The final form of the weight of feature i, denoted as fw(i) is defined by equation (8),

$$fw(i) = \frac{\sum_{j|i} P\left(f_{ij}\right) \sum_{c} P\left(c|f_{ij}\right) \log\left(\frac{P\left(c|f_{ij}\right)}{P(c)}\right)}{-Z.\sum_{j|i} P\left(f_{ij}\right) \log\left(P\left(f_{ij}\right)\right)} \quad (8)$$

Where, Z represents a constant for normalizations computed by equation (9),

$$Z = \frac{1}{n}\sum_{i} fw(i) \quad (9)$$

Where, n is training data's image count. This work normalizes fw(i) for ensuring$\sum_i fw(i) = n$. Finally, this weight value is updated to each gate in the DWLSTM classifier.

2.1.2. Recurrent neural network (RNN)

Recurrent Neural Network (RNN) is a kind of artificial neural network which is suitable for analyzing and processing image based on the weight connection between the layers. Gated Recurrent Unit (GRU) is a variant of LSTM with a gated recurrent neural network structure, and comparing with LSTM, there are two gates (update gate and reset gate) in GRU and three gates (forgetting gate, input gate, and output gate) in LSTM; for the moment, GRU has fewer training parameters than LSTM, so GRU converges quicker than LSTM during training [22, 23]. The GRU structure is shown in Fig. 4, where σand tanh are the activation functions, $c^{\langle t-1\rangle}$ is the input of the current unit, which is also the output of the previous unit, $c^{(t)}$is the output of the current unit, which links to the input of the next unit. $x^{\langle t\rangle}$ are the inputs of training data, $\hat{y}^{(t)}$is the outcome of this unit, generated by the activation function,

and represent the reset gate and the update gate, respectively, and the candidate activation $\tilde{c}^{\langle t \rangle}$ is computed

similarly to that of the traditional recurrent unit. There are two gates in GRU, one is the update gate, which preserve previous information to the current state; The value of Γ_u ranges from 0 to 1, the closer Γ_u is to zero, the more previous information it retains; the other is the reset gate, which is used to determine whether the current status and previous information are to be combined. The value of Γ_r ranges from –1 to 1, the smaller the value of Γ_r, the more previous information it ignores [24].

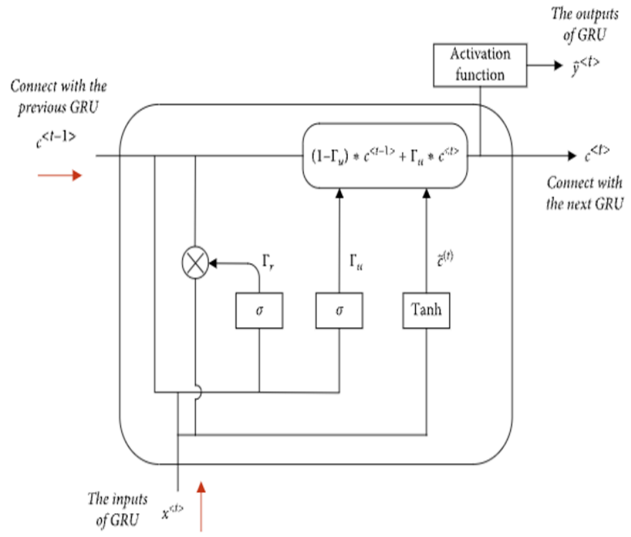


Fig. 4: A general Architecture of a single GRU cell.

According to Fig. 4, the equations of GRU can be shown by equation (10, 11, 12, and 13),

$$\Gamma_u = \sigma\left(\omega_u\left[c^{\langle t-1 \rangle}, x^{\langle t \rangle}\right] + b_u\right) \quad (10)$$

$$\Gamma_r = \sigma\left(\omega_r\left[c^{\langle t-1 \rangle}, x^{\langle t \rangle}\right] + b_r\right) \quad (11)$$

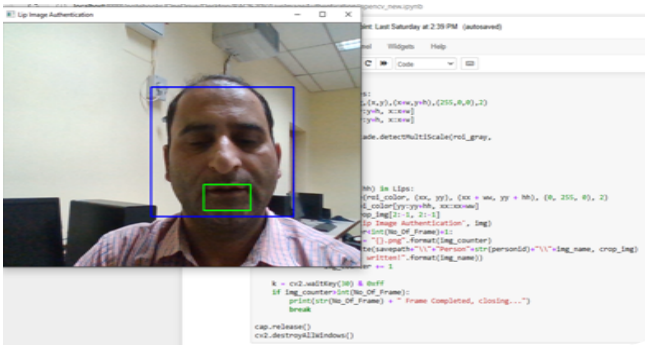
$$\tilde{c}^{\langle t \rangle} = \tanh\left(\omega_c\left[\Gamma_r * c^{\langle t-1 \rangle}, x^{\langle t \rangle}\right] + b_c\right) \quad (12)$$

$$c^{\langle t \rangle} = (1 - \Gamma_u) * c^{\langle t-1 \rangle} + \Gamma_u * \tilde{c}^{\langle t \rangle} \quad (13)$$

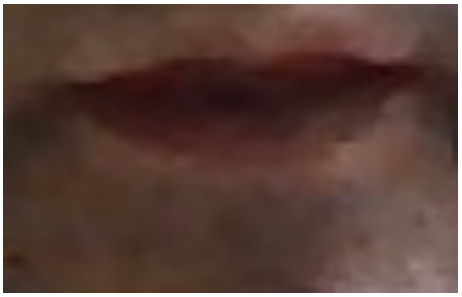
where ω_u, ω_r, and ω_c represent the training weight matrix of the update gate, the reset gate, and the candidate activation $\tilde{c}^{\langle t \rangle}$, respectively and b_u, b_r, and b_c are the bias vectors.

3 Results and Discussions

This section analyzes the performance of the proposed and existing methods using the mouth image dataset. The livemouth dataset is collected from Annamalai University Deep learning laboratory Webcam. The video is recorded using python executable code in Jupyter notebook for ten different persons. In Live Mouth Dataset, 5000 live mouth video frames are collected from 10 persons with image size of 81x48. For Each Person, 500 consecutive video frames are extracted. The dataset samples are divided into 70% fortraining and 30% for testing. The authentication methods are implemented using Anaconda with Jupyter notebook in windows. Dataset collection in laboratory environment with real time capturing of face image is shown in the Fig. 5(a). The real time captured mouth image is discussed in the Fig. 5(b).



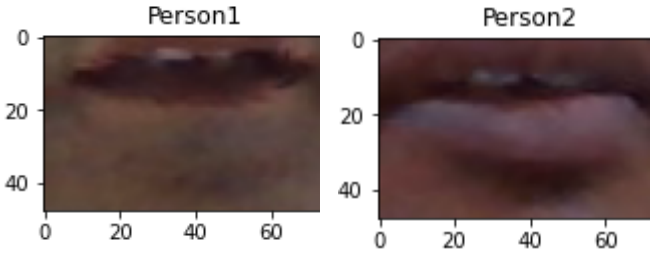
(a) Real time face image capturing



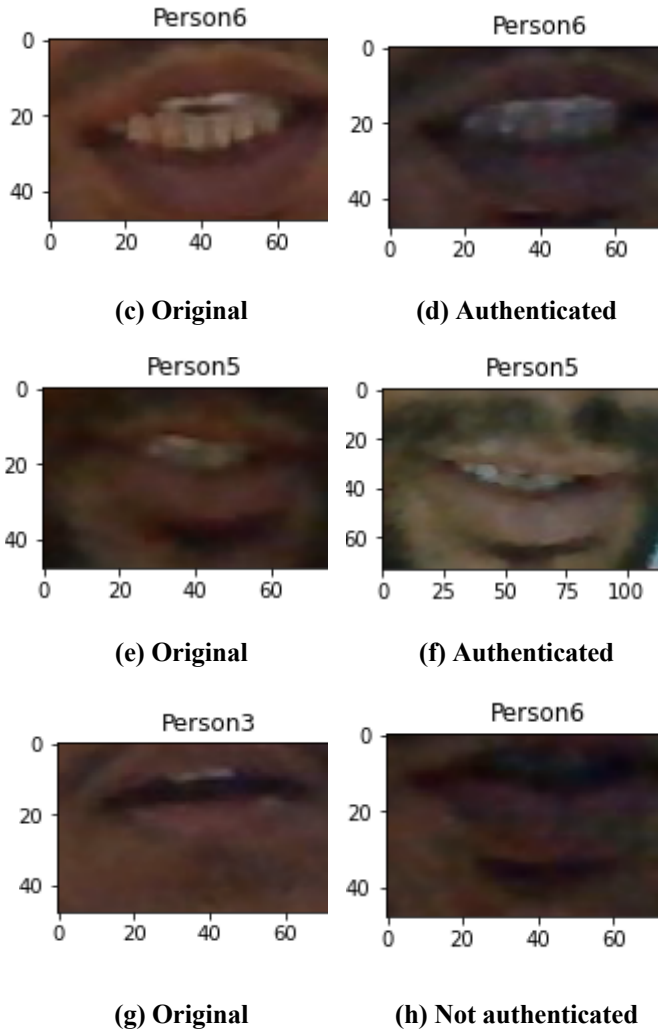
(b) Mouth image

Fig. 5: Live dataset capturing.

In the Fig. 6 shows the mouth image results of three different persons with two categories: original and authenticated/ not authenticated images. Two rows are shown in the Fig. 6, in the left column shows the four persons mouth images (Person1, Person6, Person5, and Person3) are shown Fig. 6(a), (c), (e), and (g). The second column shows the authenticated/not authenticated image results of four person mouth images (Person2, Person6, Person5, and Person6) are shown in figure 6(b), (d), (f) and (h).



(a) Original **(b) Not authenticated**



(c) Original **(d) Authenticated**

(e) Original **(f) Authenticated**

(g) Original **(h) Not authenticated**

Fig. 6: Original and authenticated/not authenticated mouth images of persons.

Code 1

```
Test_on=3
Image_to_Test_path="C:\\Users\\New
folder(2)\\LiveImageAuthentication\\testfiles\\Person6\\25.
png\\"

Mouths_Image_Authentication(Test_on=Test_on,Image_to
_Test=Image_to_Test_path)
```

Code 2

```
Test_on=6
Image_to_Test_path="C:\\Users\\New folder
(2)\\LiveImageAuthentication\\testfiles\\Person6\\30.png\\"

Mouths_Image_Authentication (Test_on=Test_on,
Image_to_Test=Image_to_Test_path)
```

Code 3

```
Test_on=1
Image_to_Test_path="C:\\Users\\New folder
(2)\\LiveImageAuthentication\\testfiles\\Person6\\50.png\\"

Mouths_Image_Authentication (Test_on=Test_on,
Image_to_Test=Image_to_Test_path)
```

Code 4

```
Test_on=5
Image_to_Test_path="C:\\Users\\New folder
(2)\\LiveImageAuthentication\\testfiles\\Person6\\30.png\\"

Mouths_Image_Authentication (Test_on=Test_on,
Image_to_Test=Image_to_Test_path)
```

The above code1, code2, code3 and code4 represent the authentication results of four different images such as Person1, Person6, Person5, and Person3. The results of the proposed system and the existing methods such as support vector machine (SVM), and probabilistic neural network (PNN) are measured using the metrics like macro precision, macro recall, macro f1-score, and accuracy. These metrics have been generated via a confusion matrix. A confusion matrix needs to be computed for each class $g_i \in G = \{1, \dots, K\}$, in such a way that the i^{th} confusion matrix assumes class $g_i$as the positive class and the remaining classes g_j with $j \neq i$ as negative class. As each confusion matrix pools together the entire observations labelled with a separate class apart from $g_i$as the negative class, this method increases the number of true negatives. This gives us:

- **True Positive (TP):** for event values that are correctly analyzed.
- **False Positive (FP):** for event values that are incorrectly analyzed.
- **True Negative (TN):** for no-event values that are correctly analyzed.
- **False Negative (FN):** for no-event values that are incorrectly analyzed.

Let us TP_i, TN_i, FP_i and $FN_i$to indicate the true positives respectively, true negatives, false negatives and false positives, in the confusion matrix associated with the i^{th}class. Let the recall here be indicated by R and precision by P. Micro average pools the performance over the least possible unit. It is computed by equation (14, 15),

$$P_{micro} = \frac{\sum_{i=1}^{|G|} TP_i}{\sum_{i=1}^{|G|} TP_i + FP_i} \quad (14)$$

$$R_{micro} = \frac{\sum_{i=1}^{|G|} TP_i}{\sum_{i=1}^{|G|} TP_i + FN_i} \quad (15)$$

The micro-averaged precision, P_{micro}, and recall, R_{micro}, give rise to the micro F1-score. It is computed by equation (16),

$$F1_{micro} = 2.\frac{P_{micro}.R_{micro}}{P_{micro} + R_{micro}} \quad (16)$$

Given that a classifier gets a large$F1_{micro}$, it denotes that it performs exceedingly well. Here, micro-average may not be sensitive to the overall predictive

performance. Due to this, the micro-average can be misleading when there is an imbalance in the class distribution.

Macro average averages over bigger groups andover the performance of individual classes than observations. It is computed by equation (17,18),

$$P_{macro} = \frac{1}{|G|}\sum_{i=1}^{|G|} TP_i / TP_i + FP_i \qquad (17)$$

$$R_{macro} = \frac{1}{|G|}\sum_{i=1}^{|G|} TP_i / TP_i + FN_i \qquad (18)$$

The recall and macro-averaged precision leads to the macro F1-score. It is computed by equation (19),

$$F1_{macro} = 2.\frac{P_{macro}.R_{macro}}{P_{macro} + R_{macro}} \qquad (19)$$

If $F1_{macro}$ has a bigger value, it points out to the fact that a classifier is able to perform well for each of the individual class. Multi-class accuracy is termed as the average of the correct predictions. It is computed by equation (20),

$$accuracy = \frac{1}{N}\sum_{k=1}^{|G|}\sum_{x:g(x)=k} I(g(x) = \hat{g}\,(x)) \qquad (20)$$

where I is defined as the indicator function, which returns 1 when there is a match between the classes and 0 otherwise. The proposed system is tested with real-time video for varying numbers of hidden layers. For authentication, 5000 images are extracted from the live webcam and it is used to evaluate the performance of the system for consecutive 'n' images (n= 1, 3, and 6). Similarly, when the model is tested for every 3 image and 6image gives 90.32% and 94.15% respectively. Among the classifiers, the proposed algorithm gives the highest validation accuracy and these models are tested in the real-time video to evaluate the performance of the system. Fig. 7 shows the confusion matrix of proposed classifier for six different person mouth images.

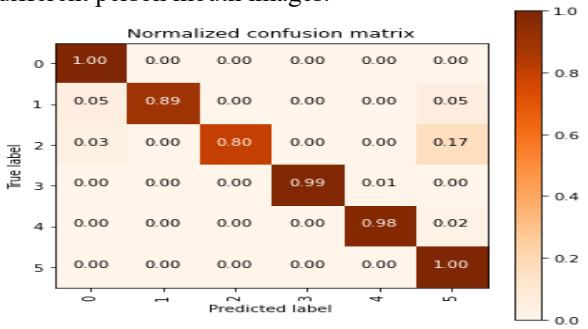


Fig. 7: Confusion matrix of proposed classifier.

Table 1: Performance of person authentication for real-time mouth images.

Methods	n=1(in %)	n=3(in %)	n=6(in %)
SVM	75.62	78.32	84.21
PNN	78.22	81.21	85.78
LSTM	82.12	85.63	90.14
GRU	84.52	86.93	92.75
DWLSTM + GRU	88.36	90.32	96.78

Performance analysis

The performance of the proposed classifier is evaluated using measures like precision, recall, F1-score, and accuracy. The precision, recall, F1-score, and accuracy for real-time testing images are shown in Table 2 for every six consecutive person images.

Table2: Performance of various person authenticationsmethods.

METHODS	METRICS (%)		
	PRECISION (%)	RECALL (%)	F1-SCORE (%)
SVM	80.12	82.72	81.42
PNN	83.79	84.81	84.30
LSTM	87.51	89.17	88.34
GRU	89.23	91.28	90.255
DWLSTM + GRU	96.83	94.33	95.33

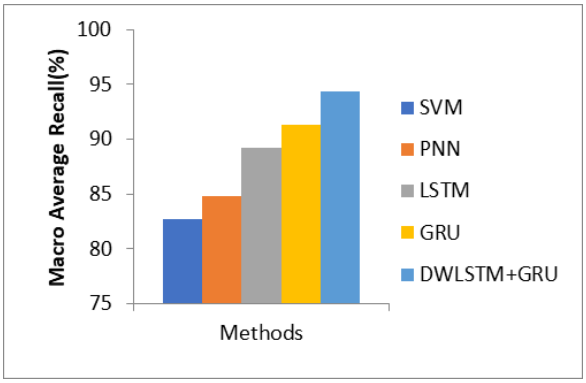


Fig. 8: Macro average precision comparison vs.

authentication methods.

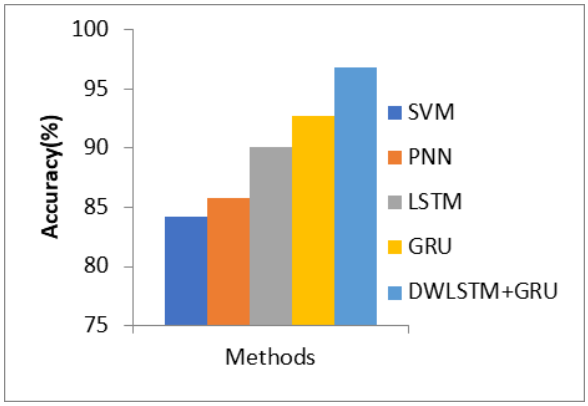


Fig. 9: Macro average recall comparison vs. authentication methods.

The macro average precision results comparison of authentication methods such as SVM, PNN, LSTM, GRU and proposed hybrid system is illustrated in figure 8. From the results it shows that the proposed hybrid system gives higher macro average precision value of 96.83%, whereas other methods such as SVM, PNN, LSTM, and GRU gives reduced value of 80.12%, 83.79%, 87.51%, and 89.23%respectively. However, the proposed algorithm has higher results than the other methods, since the proposed algorithm hybrid classifier is introduced for authentication.The classification methods like SVM, PNN, LSTM, GRU and proposed hybrid system with respect to macro average recall results are illustrated in figure 9. It shows that the proposed system gives higher macro average recall value of 94.33%, whereas other methods such as SVM, PNN, LSTM, and GRU gives reduced recall value of 82.72%, 84.81%, 89.17%, and 91.28% respectively. However, the proposed algorithm has higher results than the other methods, since the proposed algorithm hybrid classifier is introduced for authentication.

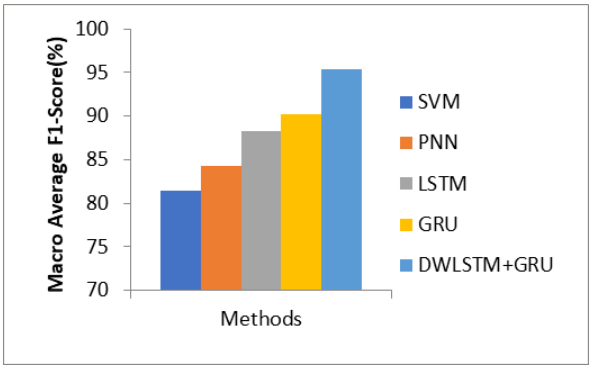


Fig.10: Performance comparison of person authentication methods (F1-Score).

The SVM, PNN, LSTM, GRU and proposed hybrid methods perform using macro average F1-Scoreare illustrated in Figure10. It shows that the proposed method gives increased F1-score value of 95.33%, whereas other methods such as SVM, PNN, LSTM, and GRU shows the performance of 81.42%, 84.30%, 88.34%, and 90.25% respectively. However, the proposed algorithm has higher results than the other methods, since the LSTM classifier is enhanced via the use of divergence function. The proposed work gives higher importance to the features by introducing the weight value to the classifier.

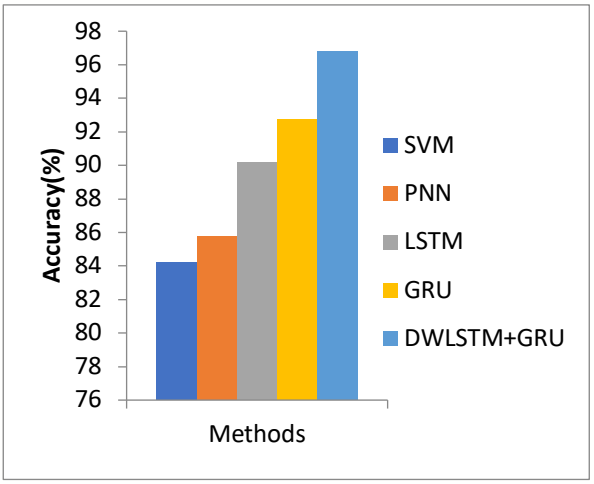


Fig.11: Performance comparison of person authentication methods (Accuracy).

The accuracy of various person authentication methods is illustrated in figure 11. It shows that the proposed method achieves a performance of 96.78%, whereas other methods such as SVM, PNN, LSTM, and GRU shows the performance of 84.21%, 85.78%, 90.14%, and 92.75% respectively.

4 Conclusions and Future Work

In this paper, a novel and effective mouth-based biometric recognition approach was introduced using hybrid divergence weight long short term memory (DWLSTM) and gated recurrent unit (GRU) namely DWLSTM+GRU. The mouth-based biometric authentication (MBBA) system is in the initial stage than the recognition of other human physical attributes such as the fingerprint, voice patterns, blood vessel patterns, or the face. The MBBA system is unique and it can be used as a universal biometric where all individuals can use it. DWLSTM+GRU system was constructed by extracting unique behavioural characteristics of users' mouths through videos on Smartphones and it is trained. To strengthen the reliability of the authentication a divergence based weight function is used in the DWLSTM. The proposed DWLSTM+GRU system gives better performance than the other classifiers for authenticating the mouths in real-time video for mobile phones. Authentication performance is measured using the metrics like precision, recall, F1-Score, and accuracy. Experiments are conducted using live mouth dataset with 5000 live mouth video frames of 10 persons. The results show that the MBBA system can achieve 96.78% accuracy.

Conflict of Interest:
There is no conflict of interest among the authors.

References

[1] Bubukayr, M.A.S. and Almaiah, M.A., 2021, *Cybersecurity concerns in smart-phones and applications: A survey*. In 2021 International Conference on Information Technology (ICIT) ., 725-731, 2021.

[2] Hui, D.O.Y., Yuen, K.K., Zahor, B.A.F.B.S.M., Wei, K.L.C. and Zaaba, Z.F., 2018, "An assessment of user authentication methods in mobile phones", *In AIP Conference Proceedings* (Vol. 2016, No. 1, p. 020116). AIP Publishing LLC.

[3] Patel, V.M., Chellappa, R., Chandra, D. and Barbello, B., 2016. Continuous user authentication on mobile devices: Recent progress and remaining challenges. *IEEE Signal Processing Magazine*., **33(4)**,49-61C, 2016.

[4] J Gunasinghe, H. and Bertino, E., 2017. PrivBioMTAuth: Privacy preserving biometrics-based and user centric protocol for user authentication from mobile phones. *IEEE Transactions on Information Forensics and Security*., **13(4)**,.1042-1057, 2017.

[5] Zhou, L., Kang, Y., Zhang, D. and Lai, J., 2016, "Harmonized authentication based on Thumb Stroke dynamics on touch screen mobile phones", *Decision Support Systems*., **92**,14-24, 2016.

[6] Wright, C. and Stewart, D., 2019, *One-shot-learning for visual lip-based biometric authentication.* In International Symposium on Visual Computing (pp. 405-417). Springer, Cham.

[7] Wright, C. and Stewart, D.W., 2020. Understanding visual lip-based biometric authentication for mobile devices. *EURASIP Journal on Information Security*., **(1)**, 1-16, 2020.

[8] Das S., K. Muhammad, S. Bakshi, I. Mukherjee, P. K. Sa, A. K. Sangaiah, A. Bruno, Lip biometric template security framework using spatial steganography, *Pattern Recognition Letters*.,**126,** 102–110, 2019.

[9] Rogowski, M., Saeed, K., Rybnik, M., Tabedzki, M. and Adamski, M., 2013, *User authentication for mobile devices*. In IFIP International Conference on Computer Information Systems and Industrial Management (pp. 47-58). Springer, Berlin, Heidelberg.

[10] Zhou, L., Kang, Y., Zhang, D. and Lai, J., 2016. Harmonized authentication based on Thumb Stroke dynamics on touch screen mobile phones. *Decision Support Systems*., **92**, 14-24, 2016.

[11] Raman R., P. K. Sa, B. Majhi, S. Bakshi, Fusion of shape and texture features for lip biometry in mobile devices, *Mobile Biometrics*.,**3**, 1-13.

[12] Wrobel, K., Doroz, R., Porwik, P., Naruniec, J. and Kowalski, M., 2017. Using a probabilistic neural network for lip-based biometric verification. *Engineering Applications of Artificial Intelligence*., **64**, 112-127, 2017.

[13] Cruz, J.C.D., Garcia, R.G., Go, S.M.M., Regala, J.L. and Yano, L.J.T., 2020, *Lip Biometric Authentication Using Viola-Jones and Appearance Based Model (AAM) System*. In 2020 35th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC) ., 372-377.

[14] Lu, L., Yu, J., Chen, Y., Liu, H., Zhu, Y., Liu, Y. and Li, M., 2018, Lippass: *Lip reading-based user authentication on smartphones leveraging acoustic signals*. In IEEE INFOCOM 2018-IEEE Conference on Computer Communications., 1466-1474, 2018.

[15] Wrobel, K., Doroz, R., Porwik, P. and Bernas, M., 2018. Personal identification utilizing lip print furrow based patterns. A new approach. *Pattern Recognition*., **81**, 585-600, 2018.

[16] Porwik, P., Doroz, R. and Wrobel, K., 2019. An ensemble learning approach to lip-based biometric verification, with a dynamic selection of classifiers. *Expert Systems with Applications*., **115**, 673-68, 2019.

[17] Wright C., D. Stewart, P. Miller, F. Campbell-West, R. Dahyot, G. Lacey, K. Dawson-Howe, F. Pitie´e´, D. Moloney, *Investigation into DCT feature selection for visual lip-based biometric authentication*. In Irish Machine Vision & Image Processing Conference proceedings., 11–18.

[18] Zhu, X., Sobihani, P. and Guo, H., 2015,.*Long short-term memory over recursive structures*. In International Conference on Machine Learning (pp. 1604-1612). PMLR.

[19] Zhang, K., Chao, W.L., Sha, F. and Grauman, K., 2016, *Video summarization with long short-term memory*. In European conference on computer vision (pp. 766-782). Springer, Cham.

[20] Mateus, B.C.; Mendes, M.; Farinha, J.T.; Cardoso, A.M. Anticipating Future Behavior of an Industrial Press Using LSTM Networks. *Appl. Sci*., **11**, 1-16, 2021.

[21] Sagheer, A.; Hamdoun, H.; Youness, H. Deep LSTM-Based Transfer Learning Approach for Coherent Forecasts in Hierarchical Time Series. *Sensors*., **21**, 1-23, 2021.

[22] Dey, P., Hossain, E., Hossain, M., Chowdhury, M.A., Alam, M., Hossain, M.S. and Andersson, K., 2021. Comparative Analysis of Recurrent Neural Networks in Stock Price Prediction for Different Frequency Domains. *Algorithms*., **14(8)**, 1-20, 2021.

[23] Gao, X., Li, X., Zhao, B., Ji, W., Jing, X. and He, Y., 2019. Short-term electricity load forecasting model based on EMD-GRU with feature selection. *Energies*, **12(6)**, pp.1-18.

[24] Islam, M.S. and Hossain, E., 2020. Foreign exchange currency rate prediction using a GRU-LSTM Hybrid *Network. Soft Computing Letters*., **1-20**, 2020.