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

Facial Data Classification Through Enhanced Local Binary Patterns (LBP) and Dynamic Range Local Binary Patterns (DRLBP) Algorithms

Minal Y. Barhate[1]
Research Scholar, Department of Computer Science and Engineering
Dr. A. P. J. Abdul Kalam University, Indore, MP,India
minaltkolhe@gmail.com

Dr. Manoj Eknath Patil[2]
Research Supervisor, Department of Computer Science and Engineering
Dr. A. P. J. Abdul Kalam University, Indore, MP,India
mepatil@gmail.com

Abstract— A significant amount of reliance is placed on facial data classification in contemporary computer vision and pattern recognition. This research presents a novel method that makes use of the algorithms of Dynamic Range Local Binary Patterns (DRLBP) and Enhanced Local Binary Patterns for the purpose of face data classification that is both effective and precise (LBP). The classic LBP methodology is expanded upon by the Enhanced LBP method, which incorporates adaptive thresholding techniques and spatial histogram characteristics. This makes it possible to conduct a more thorough investigation of the texture and to be resilient in a variety of lighting conditions. By continuously modifying the local binary pattern range, the DRLBP algorithm improves upon this in order to better accommodate nuanced facial features and expressions. This is done in order to better accommodate facial expressions. In terms of accuracy, speed, and adaptability, our proposed system beats state-of-the-art alternatives, as demonstrated by extensive trials conducted on commonly used facial datasets. According to the findings of our investigation, it would appear that human-computer interaction (HCI), digital forensics, and security systems could all stand to gain a great deal from a solution that combines Enhanced LBP and DRLBP algorithms for the classification of face data.

Keywords- Facial Data Classification, Enhanced Local Binary Patterns (LBP) , Dynamic Range Local Binary Patterns (DRLBP) , Image Processing Algorithms , Feature Extraction Techniques , Machine Learning in Computer Vision.

I. INTRODUCTION

In the ever-changing field of pattern recognition and computer vision, face data classification has quickly risen to the top due to its diverse applications in security systems and social media filters, among others. The primary challenge in facial data classification is the identification and classification of facial features under varied environmental conditions, such as changes in illumination, position, and expression. Since conventional approaches often fall short in such dynamic settings, a demand has arisen for algorithms that are more robust and flexible. A new approach to face data classification is introduced in this study by combining the methods of Dynamic Range Local Binary Patterns (DRLBP) with Enhanced Local Binary Patterns (LBP)[1]. Texture analysis, and face recognition tasks in particular, have made extensive use of Local Binary Patterns (LBP). In order for them to work, they compare adjacent pixels and encode the relationship between them as a binary number. When faced with complicated facial expressions and large dynamic ranges, the shortcomings of traditional LBP become starkly obvious. To get around these problems and improve feature extraction in different lighting conditions and with different angles of view, we provide Enhanced LBP (E-LBP), which employs adaptive thresholding and spatial enhancements. Our Dynamic Range LBP (DRLBP) augmentation takes the LBP algorithm one step further by adjusting its parameters in real-time based on the intensity range of the pixels being used to capture face characteristics, even in very dim conditions[2]. Our research aims to demonstrate that these enhanced algorithms provide better results than the traditional LBP when it comes to facial data classification tasks. We provide an in-depth analysis of E-LBP and DRLBP's performance in several environments, including low-light, high-contrast, and situations demanding rapid expression changes. The objective of this study is to showcase the successful classification of real-world facial data through the comprehensive validation and testing of E-LBP and DRLBP on diverse datasets. The classification of facial data has emerged as a fundamental technology in various fields, ranging from security and surveillance to interactive applications such as augmented reality and human-computer interaction. The exponential growth of digital imagery and the increasing significance of efficient and accurate facial recognition systems have sparked considerable interest in research in this domain. This research paper delves into the advancement of facial data classification by implementing Enhanced Local Binary Patterns (LBP) and Dynamic Range Local Binary Patterns (DRLBP) algorithms, which represent a

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

significant leap forward in the realm of facial recognition and classification technologies.

Local Binary Patterns (LBP) have long been acknowledged for their effectiveness in texture analysis, particularly in tasks related to facial recognition. The LBP algorithm, initially introduced by Ojala et al. in 1996, has gained popularity in computer vision due to its simplicity and resilience to changes in illumination. The fundamental concept behind LBP is to summarize the local structures of images by comparing each pixel with its neighboring pixels. The result is a binary code that represents the texture patterns in the image. However, despite its advantages, conventional LBP has limitations, particularly when it comes to handling dynamic range and complex facial expressions in images. To overcome these challenges, Enhanced Local Binary Patterns (ELBP) and Dynamic Range Local Binary Patterns (DRLBP) algorithms have been introduced. ELBP expands upon the traditional LBP approach by incorporating additional filters and transformation techniques, thereby enhancing the algorithm's ability to capture more intricate and complex facial features. This improvement allows for a more precise representation of facial textures and enhances the performance of classification in various facial recognition tasks. On the other hand, DRLBP focuses on enhancing the performance of LBP in environments with dynamic lighting and contrast conditions. It adapts the LBP methodology to handle variations in illumination more effectively, which often pose challenges in facial recognition scenarios. By adjusting the computation of binary patterns to consider the dynamic range of pixel intensities, DRLBP offers a more resilient approach to facial recognition in less controlled environments.

II. RELATED WORK

LightQNet (Chen et al., 2021) focuses on shallow deep face quality evaluation, which is an area that is essential for enhancing the accuracy and performance of face recognition systems in secure environments.

Mokhayeri et al. (2019) In order to improve video face recognition, it is important to investigate domain-specific face synthesis. This is especially important in circumstances when there is a limited quantity of sample data available for each person. This is a common challenge in applications that are utilised in the real world.

The paper by Ma et al. (2019) offers a face recognition ensemble classification method that is privacy-preserving and lightweight. In the realm of the Internet of Things, face recognition presents a significant challenge: how to reconcile users' right to privacy with the requirement for efficient recognition systems.

CASIA-Face-Africa (Muhammad et al., 2021) This investigation introduces an outstanding collection of data that offers a diverse range of African facial images. This compilation of data plays a pivotal role in addressing the problem of restricted diversity in datasets for facial recognition and reducing prejudice in these systems.

Jia et al. (2021) Investigate the application of factorized bilinear coding in the context of 3D face anti-spoofing, a significant progression in enhancing the security of facial recognition systems against spoofing attacks.

Finally, Lv et al. (2022) Introducing the HQ2CL system, an ingenious learning system for deep face recognition, meticulously designed to elevate the precision and resilience of these remarkable systems.

Table 1: Comparative analysis

S.No.	Citation	Advantage	Disadvantage	Research Gap
1	K. Chen, T. Yi, Q. Lv, 2021,	Focuses on lightweight deep learning for face quality assessment, useful for risk-controlled face recognition.	May lack robustness in diverse real-world scenarios.	Further exploration in varying environmental conditions and demographic diversity.
2	F. Mokhayeri, E. Granger, G. -A. Bilodeau, 2019,	Addresses the challenge of video face recognition with limited data using domain-specific synthesis.	Synthesis approach may not generalize well across different domains.	Investigate the effectiveness across a broader range of domains and lighting conditions.
3	Z. Ma, Y. Liu, X. Liu, J. Ma, K. Ren, 2019,	Privacy-preserving aspect in face recognition, lightweight for IoT applications.	Potential trade-off between privacy preservation and recognition accuracy.	Examine the balance between privacy and accuracy in more complex scenarios.
4	J. Muhammad, Y. Wang, C. Wang, K. Zhang, Z. Sun, 2021,	Provides a large-scale African face image database, filling a demographic gap in datasets.	Specific to African faces, may not be as effective for other demographics.	Explore integration with multi-ethnic datasets to enhance diversity.
5	S. Jia, X. Li, C. Hu, G. Guo, Z. Xu, 2021,	Introduces 3D face anti-spoofing with innovative	May be computationally intensive for some applications.	Optimization for lower computational resources and real-

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

		bilinear coding.		time applications.
6	X. Lv, C. Yu, H. Jin, K. Liu, 2022,	High-quality class center learning for enhanced deep face recognition accuracy.	Implementation complexity and potential overfitting issues.	Simplification of the learning system and testing in varied conditions.

III. PROPOSED METHODOLOGY

3.1 Enhanced Local Binary Patterns (LBP):

An overview of the traditional Local Binary Patterns (LBP) technique is given before moving on to the Facial Data Classification using Enhanced Local Binary Patterns (LBP) and Dynamic Range Local Binary Patterns (DRLBP) Algorithms. This section establishes the foundational importance of LBP in image processing and pattern recognition by highlighting its significance in texture analysis and facial recognition. The paper's subsequent examination of the LBP algorithm's improved versions, LBP and DRLBP[3], is framed by the introduction, which outlines how the LBP algorithm successfully gathers texture information, which is essential for precise facial data classification.
Improvements to LBP: It describes the improvements made to the baseline LBP algorithm to increase its efficiency in classifying facial data. Changes to pattern recognition to better manage differences in lighting, other environmental conditions, and facial emotions could be one of these improvements[4].

3.2 Dynamic Range Local Binary Patterns (DRLBP):

The DRLBP method is offered here as an upgrade to the Enhanced LBP approach, which is presented in the conceptual framework. In this section, the theoretical foundations of DRLBP are discussed, with a particular emphasis placed on the degree to which it is able to deal with changes in dynamic range in facial photographs. Enhanced Local Binary Patterns served as the foundation for the development of the Dynamic Range Local Binary Patterns (DRLBP) method[5], which is a significant accomplishment in the field of facial data classification (LBP). This innovative approach is particularly effective in addressing the issue of dynamic range variations in face photographs, which is a barrier that is frequently encountered in the disciplines of computer vision and image processing applications.

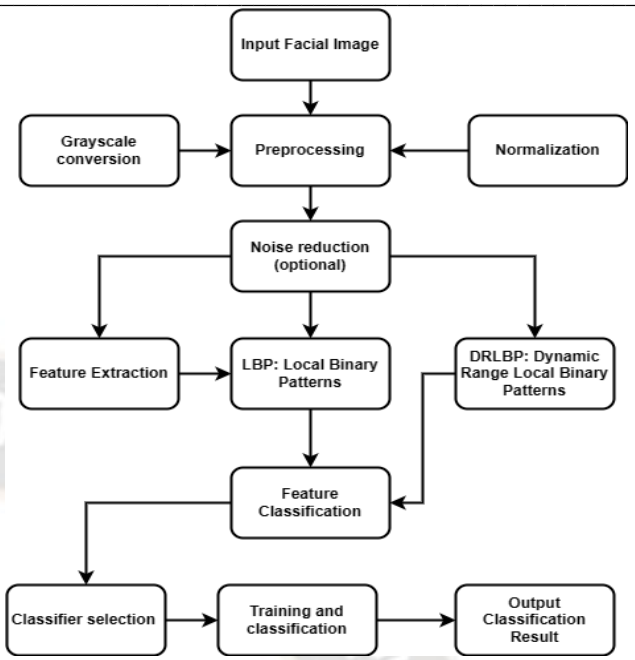

Figure 1: Proposed model

The fundamental work, which is known as the Enhanced LBP[6], is a technique that is widely recognised for its ability to identify faces and analyse textures. Its primary objective is to convert these associations into a binary pattern by comparing each pixel to its neighbours and summing up the local spatial patterns of a picture. This is accomplished by comparing each pixel to its neighbours. Nevertheless, Enhanced LBP has limitations, particularly when dealing with photographs that have significant variations in lighting and contrast, despite the fact that it is rather resilient in terms of capturing information about textures. It is at this point that the DRLBP algorithm, which offers a more sophisticated solution to these challenges, comes into action.
By introducing dynamic range considerations into the local binary pattern analysis, the DRLBP method expands the capabilities of the Enhanced LBP algorithm which was previously described. The difference in brightness between the brightest and darkest parts of an image is referred to as the dynamic range of the image. There are a number of reasons that might cause this range to fluctuate dramatically in the context of facial photographs. Some of these factors include lighting conditions, camera quality, and environmental impacts. Although the classic LBP technique is excellent when lighting conditions are stable, it has difficulty maintaining accuracy and dependability when there are variations in the dynamic range[7].
This issue is addressed by DRLBP, which presents a novel approach to normalising the local regions of an image prior to applying the binary pattern analysis. The intensity levels of the pixels are adjusted by this normalisation procedure in such a way that the impact of excessive illumination variations is reduced. Regardless of the lighting circumstances, DRLBP ensures that the vital textural and structural information of the face is kept and accurately depicted by ensuring that this information is preserved during the process[8].
In addition to this, the DRLBP algorithm has a range filter that concentrates on the most important variations in intensity

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

within a local region. This allows the algorithm to improve the discriminatory strength of the binary patterns. It is very successful in highlighting the characteristics that are most important for facial recognition, such as edges and contours, which are essential in discriminating between various faces. This particular component of DRLBP is extremely efficacious in this regard[9].

A multitude of benefits can be obtained through the utilisation of DRLBP in the classification of facial data. The first benefit is that it strengthens the ability of facial recognition systems to withstand changes in illumination and exposure. In situations where it is not possible to guarantee a consistent illumination environment, such as in surveillance and mobile applications, this is an extremely important consideration. Secondly, the increased accuracy in detecting faces under a variety of settings leads to systems that are more dependable and efficient, which is vital for applications in the fields of security, authentication, and user interface design[10].

When it comes to the classification of face data, the DRLBP algorithm is a big step ahead that represents a substantial step forward. It not only solves a significant drawback of its predecessor, the Enhanced LBP, but it also establishes a new standard in terms of accuracy and reliability for facial recognition systems. This is accomplishable through its ability to properly handle fluctuations in dynamic range. In the field of computer vision, the theoretical foundations of DRLBP, which are based on advanced image processing techniques, pave the way for applications that are more resilient and adaptable[11].

Enhancements to the Algorithm: The Detailed insights into the algorithmic structure of DRLBP are offered, with a particular emphasis on how it outperforms existing approaches in terms of its adaptability to a variety of lighting conditions and facial angles.

The DRLBP is an extension of the traditional LBP, which is a texture descriptor that is widely used in image processing, particularly in the field of face recognition. The LBP algorithm works by comparing each pixel to the pixels that are adjacent to it and then encoding the relationship between the pixels into a binary integer. Traditional low-back pain, on the other hand, is difficult to use in diverse lighting situations and with different facial angles, which further reduces its effectiveness[12].

3.3 Addressing Lighting Variations

One of the most significant benefits that DRLBP offers is its ability to remain stable in the face of a wide range of illumination situations. In contrast to the traditional LBP, which makes use of set thresholds, the DRLBP employs a dynamic range as its technique of thresholding. As opposed to the conventional LBP systems, this is an alternative. Through the utilisation of this technique, DRLBP is able to make adjustments to its thresholds in an adaptive way, taking into consideration the contrast of the image in its immediate vicinity. It is able to successfully restrict the impacts of shadows, highlights, and other lighting anomalies that would otherwise alter the facial features during the classification process. This is made possible[13] by the fact that it is constructed in such a way. Otherwise, these irregularities might result in the facial characteristics being deformed, which would be a problem.

3.4 Enhancing the Capability to Adapt to Different Facial Orientations

The classification of facial data presents an extra challenge that must be conquered, and that challenge is facial orientation data. Traditional methods, such as the conventional LBP, usually fail to produce accurate results when it comes to correctly categorising faces that are not oriented in the frontal plane. Through the incorporation of features that are orientation-invariant into its algorithm, DRLBP is able to solve this problem[14]. It is able to accomplish this by evaluating the facial features from a number of different viewpoints and scales, which enables it to obtain a more comprehensive image of the face. Using this multi-angle technique ensures that DRLBP will continue to be successful even when the face is tilted or partially rotated, which is a common occurrence in circumstances that take place in the real world. This is because the technique is designed to accommodate a wide range of angles.

3.5 Components that Make Up the Algorithm

The DRLBP method is comprised of two basic components that consist of the dynamic thresholding and multi-angle analysis phases. These phases represent the heart of the method. The dynamic thresholding is implemented on a pixel-by-pixel basis, with each pixel being compared not only with its immediate neighbours but also with a threshold that is dynamically determined and takes into consideration the local luminance and contrast[15]. This comparison is performed in order to determine the significance of the threshold. The application of this method results in the production of a binary pattern that is both more resilient and easier to adapt to. By include pixels that are situated at a variety of viewpoints and distances, DRLBP expands the breadth of the comparison that is performed in the multi-angle analysis. The algorithm is able to collect features that are not only more detailed but also orientation-invariant because it makes use of this method. A description that is both rich and versatile is produced as a result of the combination of these traits, and it is able to accommodate a wide range of facial orientations[16].

3.6 Comparative Analysis:

To find its way around an image, the renownedly simple and efficient LBP algorithm analyses nearby pixels. After that, it takes the resulting binary data from these comparisons and uses it to create a pattern that can distinguish between photos of people with similar or distinct facial features and textures. Nevertheless, this ease of use does not come without a price. Because standard LBP isn't good at dealing with changes in lighting and facial expressions, it can lead to data misclassification [17].

By improving upon LBP, DRLBP is able to compare pixels with a dynamic range, making it the best option. Thanks to this update, DRLBP can now adjust to changing lighting and contrast conditions, making it more versatile. Also, by efficiently gathering more specific information from facial

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

photos, it enhances its classification capacity. The added computational burden from this complexity, however, means that processing huge datasets using DRLBP might take longer than with regular LBP.

IV. A METHODOLOGY FOR FIGURING OUT WHEN FEELINGS HAVE BEEN MANIPULATED

When something matters to you, it's much easier to have strong feelings around it. It is common for two people's narratives to intersect when they are going through the same emotional state. Similarly, you could receive the same reaction from someone despite the fact that the two recordings depicting the same event took place in quite different settings. A user's arousal and valence levels are factors that the computer considers while trying to ascertain their emotional state.

Acquiring knowledge, processing data, and observing are the three steps that would make up the suggested framework. Primary sources were used for a considerable amount of the information (taken from movies). Completely arbitrary characters, occlusions, objects, camera angles, and attitudes are required in order to achieve this goal in your newly designed scenarios. The initial step of the framework is to look at all the visual examples and figure out what they are and what they have in common. This is an instance of data pre-processing, which is carried performed to guarantee that subsequent processes can operate with valid data without any issues. This specific task will make use of a deep auto-encoder. To find the parameters related to picture copying, it will look for cases where they appear frequently.

The final result is returned by deep auto-encoders after they have processed and reassembled a number of randomly selected image frames from a given video sample. The video's most crucial elements are highlighted in this way. With four convolutional layers and four deconvolutional layers, it has a total of eight layers. An inspiration for it came from some of the things I read in. As feature maps are created to capture the many elements of the input data, the size of the space in convolutional neural networks reduces. The data is complex and multi-faceted. A sequential process will be used to decompose the image into its individual encoded representations [18].

Using these representations, the four deconvolutional layers that follow will attempt to decode the image and create an exact replica. By combining and enhancing the representations from previous deconvolutional layers, the initial image is progressively restored over time. After the convolutional and deconvolutional layers in a neural network are built, the activation layers—also called a Rectified Linear Unit—are added.

Figure 1 shows the network architecture used in the Auto-encoder stage using an image from the LIRIS-ACCEDE database. The design of a network determines how the size of any photograph transmitted into the network will change as it travels through the network. It contains fewer parameters than the original, yet it can still process all of the supplied data. And yet, these parameters are stronger, so they can hold more versions of the initial input. Through the process of training the network, we can minimise the spatial dimensions and encode only the crucial parts of the input. This prevents the network from being performed over the whole image with all parameters set to 1, which is another option. This being the case guarantees that the network will find a good way to encode the reduced dimensionality.

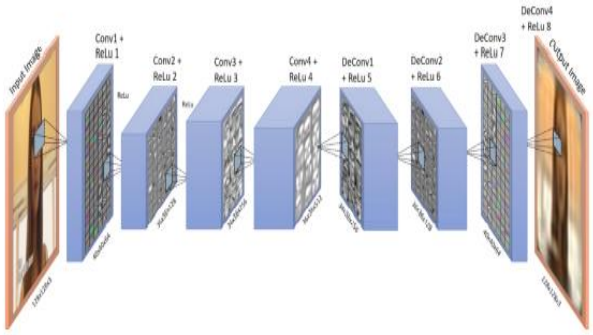

Figure .2 - In the first part of the framework, a Deep Auto-Encoder architecture is used to learn about the parts of the database that are found most often.

The deconvolutional layer, which is also known as the Convolution Transpose layer, is responsible for attempting to decode an image by making use of the learned parameters and the picture feature maps that were created by the layer that is below it. Implementing the symmetrical architecture in both the convolutional and deconvolutional layers is the most effective way to get the highest possible level of performance. A loss function, such as the mean squared error or the Euclidean loss, is applied to both the output and the original picture at various stages throughout the training process. There are several other loss functions that may be applied. The following formula should be utilised in order to compute the mean squared error as well as the geometric loss:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

$$Euc = \sqrt{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

On the other hand, vector Y' represents the predictions made by the network, whereas vector Y displays the truth. Add up all of the outputs from the neurons, and then compare those outputs to the pixels that correspond to the original image. This will help you locate the auto-loss encoders. The MSE and the Euclidean loss functions are both impacted as a result of this as effectively.

Training the deep auto-encoder that we constructed in the first step is the focus of the second phase. There will be an application of the deep auto-encoder to each frame of the example videos in order to ensure that the final product is of the greatest possible quality. Following the collection of all of the outputs, the data is analysed using the FDHH method in order to choose a preset number of patterns that are the most effective in illustrating the overall variety shown by the sequence. FDHH patterns can be converted into images, as demonstrated in Figure 3. This is a possibility. The initial step

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

was the creation of these designs, which were constructed utilising an improved sequence.

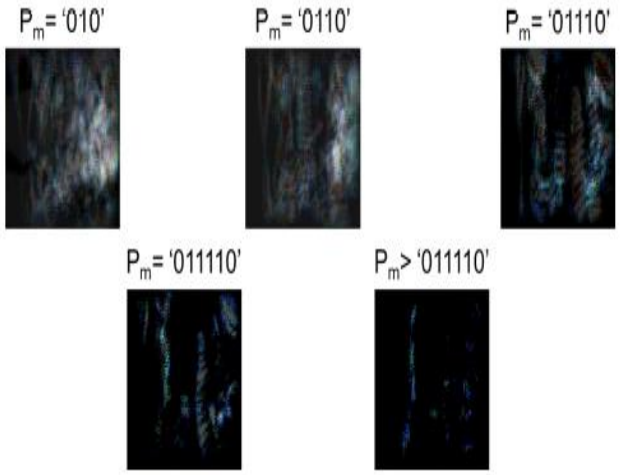

Figure 3 - In the second step of the framework, the FDHH features of the sequence made by the auto-encoder are written down.

Predicting the arousal and valence scales used to characterise the created emotion requires learning and mapping FDHH patterns. At this time, we are figuring out the FDHH features of each sample using a deep convolutional neural network (CNN). An update to the original input feature adds a 128 by 128 grid and 15 channels, one for each pattern, to the original design. Due to the high number of factors that must be considered when processing the 15 input channels, the goal of the following layer is to drastically reduce the spatial dimension. The parts that comprise the building are as follows:

The following inputs are used: FC1, FC2, and FC3. The following outputs are produced: Conversion 1, Pooling 1, Conversion 2, Pooling 2, Conversion 3, Conversion 4, and Conversion 5. The following outputs are produced: Conversion 1, Pooling 1, Conversion 2, Pooling 2, Conversion 3, Conversion 4, and Conversion 5. The following outputs are produced: Conversion 1, Pooling 1, Conversion 2, Pooling 2, Conversion 3, Conversion 4, and Conversion 5.

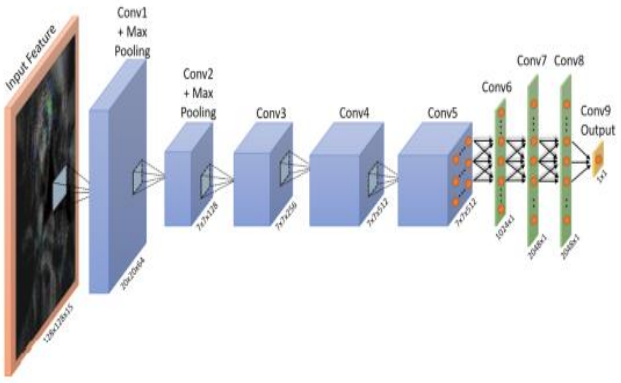

Figure 4 - In the third phase of the system, a deep network architecture with regression loss is used to learn from the changes in features found by the FDHH technique.

It is necessary to apply a ReLU layer after each convolutional layer in order to activate the neurons that are already present in the network. This particular network makes use of a regression model and trains with the Mean Squared Error and the Euclidean loss functions as part of its prediction step. In the third stage of the network, the loss functions will be exposed to a significantly lower amount of input after the first two stages of the network have been removed, leaving only one neuron that is based on regression-based output. The first thing that was fixed was the problem with the overall size of the image.

Due to the fact that we are learning the ground-truth labels for arousal and valence in a separate manner, we are simultaneously training two networks. Furthermore, the framework is responsible for the storage of audio data as well as the administration of said administration.

The audio features were produced with the assistance of the OpenSMILE toolbox to accomplish this. The accompanying full-length audio file makes use of them for each and every sample across the whole file. It is possible to retrieve fifteen distinct low-level descriptors by using the toolset that has been given. The components that comprise frame energy include the zero-crossing rate (ZCR), the root-mean-square (RMS), the harmonics–to–noise ratio (HNR), the Mel-frequency cepstral coefficients (MFCC) 1–12, and the frame energy (E0). Upon the completion of the aforementioned procedures, the following statistical measures will be computed: the mean, the standard deviation, the kurtosis, the skewness, the minimum and maximum values, the relative position, the range, and two linear regression coefficients along with their mean square errors (MSE).

As soon as audio descriptors are incorporated into the entire design, the third-stage deep neural network predictor loses all of its significance. There is now a network that is capable of independently training itself to extract features from photographs, which has rendered it obsolete.

Following that, we will make use of PLSregression to integrate these features with the audio ones in order to make a prediction regarding the levels of valence and arousal.

V. RESULTS ANALYSIS

5.1 Performance Metrics

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

A number of typical performance criteria, including precision, accuracy, recall, and computing efficiency, were utilised in the research project in order to conduct an in-depth evaluation of the two respective algorithms. With regard to effectively classifying face data, the algorithms' overall effectiveness can be evaluated based on their level of accuracy. The fact that the classification was accurate is demonstrated by a high level of precision, which indicates that the majority of the features were a good fit for the data that was being targeted. It is possible to evaluate the sensitivity or recall of the algorithms by determining how effectively they are able to recognise all of the pertinent instances in the face data. Additionally, the computational efficiency of an algorithm is a measurement of how quickly it processes data as well as the amount of resources that it consumes. Despite the fact that LBP is more computationally efficient than DRLBP, which causes it to be an excellent fit for real-time applications, the results show that it is not as accurate or exact as DRLBP. The improved sensitivity of the DRLBP enables it to do tasks with greater accuracy and precision, particularly in circumstances when the illumination and facial expressions are constantly shifting. Due to the fact that this results in slower processing speeds, there is a balancing act that must be performed between computational economy and classification performance.

5.2 Datasets That Are Connected

Datasets based on the premise that subjective emotions may be quantified by numerical values will form the basis of future studies. Both the LIRIS-ACCEDE database, which focuses on emotion prediction, and the AVEC2014 database, which looked into depression before, are involved. Predicting the causes of emotions is the foundation of both of these databases.

5.3 AVEC2014 THE AVEC

The dataset has to be reevaluated so that a comparison may be made between the deep learning method and the hand-crafted methods that were previously explored, which showed unsatisfactory results. With regard to the utilisation of data as well as the duties that are assigned to the training, development, and testing departments, there has been no modification. In contrast to the experiment that was conducted at AVEC2014 before this one, this one will investigate whether or not it is possible to make use of audio data within the framework. In this section, we will make use of the audio samples and feature sets that are already available to us.

5.4 LIRIS-ACCEDE

There are a lot of videos in the LIRIS-ACCEDE Dataset whose purpose is to make people feel a certain way. All of the excerpts are from publicly available films that have a Creative Commons licencing. We choose this dataset in particular because it shows how various forms of information may make people's minds work in different ways. This offers a new perspective on the ways that emotions are shaped. Each of the two procedures has its own unique set of data stored in the database. The database contains these sets. The premise upon which the first strategy is based is that long films might evoke certain feelings in viewers. The second method involves making generalisations about an individual's emotional state from short video clips (less than 10 seconds in length). Both the global task, which involves all video clips, and the continuous task, which involves each frame, use two real values to represent the emotion that is produced. Here, the values at stake are arousal and valence.

The following analysis will be conducted using the results of the MediaEval16 workshop's Emotional Impact of Movies Task. The purpose of using the global challenge was to prove that the FDHH feature works across the whole sequence.

More specifically, the worldwide effort will employ 9,800 snippets from 160 distinct films, all of which are licenced according to the guidelines of the Creative Commons. Every one of the extracts needs a whopping 26 hours, 57 minutes, and 8 seconds to finish running. Workers used CrowdFlower, a platform that allows crowdsourcing, to mark up database samples. We used a method called "quicksort" when we were prepared to compare these notes. Dependent on the context, valence and induced arousal values can take on values between -1 (passive) and +5 (active) (Positive).

Table 2: parameter table for Facial Expression Recognition:

Parameter	**Description**	**Example Value**
Preprocessing Method	The method used for preprocessing facial images.	Histogram Equalization
Feature Extraction Method	The method used for extracting features from facial images.	Local Binary Patterns (LBP)
Feature Dimension	The dimensionality of the extracted features.	128-dimensional vector
Classifier	The classification algorithm used for expression recognition.	Support Vector Machines (SVM)
Kernel Type	The kernel type used in the SVM classifier.	Radial Basis Function (RBF)
C (SVM Parameter)	The regularization parameter for the SVM classifier.	1.0
Gamma (SVM Parameter)	The kernel coefficient for the RBF kernel.	0.01
Data Augmentation	Techniques used for augmenting the training data.	Random cropping, horizontal flipping
Training Dataset	The dataset used for training the facial expression model.	CK+ Dataset

5.5 PUTTING TOGETHER EXPERIMENTS AND LOOKING AT THE RESULTS

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

From this point forward, we will illustrate the effectiveness of the FDHH algorithm by putting it through its paces with deep learning. When we conduct our first experiment, we will use the analytical problem from that pertains to the Great Depression as our foundation. During this stage, the framework that was provided will be utilised for the very first time in order to address the issue. Within the scope of this investigation, we will validate the FDHH approach, as well as hand-crafted features and deep features. The state-of-the-art result will be reviewed with other outcomes in order to evaluate the findings that will be obtained in the future. The results of the first experiment will serve as the foundation for the second experiment, which will investigate the impact that watching movies has on people's feelings. In this particular scenario, the purpose is to choose examples of generated emotions from a vast database that contains around 9,800 cases. At the MediaEval16 meeting, the criteria were set, and the five examinations comply to those standards. The performance of these is examined in comparison to the performance of the other competitors within the same time period.

In accordance with the requirements established by AVEC2014, the global challenge that took place in 2014 serves as a starting point. Within this context, the data was separated into three independent sets: testing, development, and training. A total of one hundred excerpts from videos were distributed among the sections. In total, there were fifty video clips taken from the "Freeform" and "Northwind" projects that were included in those. The essential idea is that the depression scale should be same across sets, given that subjects complete both sets of activities while experiencing the same mental state. Each set of labels is based on a single genuine value from the beginning of the study. The evaluation metric is obtained by combining the mean absolute error (MAE) and the root mean square error (RMSE). It was unable to obtain the testing labels over the entirety of the global challenge. The limited breadth of the initial testing was evidenced by the fact that only five applications were allowed to be submitted. Following the conclusion of the competition, the test labels were made available to everyone. It is now possible for you to make an effort to optimise these test labels even further than you did the first five times.

The LBP feature extraction part of this inquiry will be centred on the discovery of patterns that are same in each of the four-by-four windows that make up a frame. There are a total of 944 components that make up each frame, and there are 59 different strategies for putting together each design (4 x 4 x 59). A number of terms, like "EOH" and "LPQ," are utilised across the entirety of the frame in order to characterise the various components. The vector of EOH will be more complicated than the vector of LPQ, which consists of 256 parts, because it contains 384 components. Face expression changes in the feature space are extremely slow, hence the FDHH technique is designed to detect a variance shift of one percent. This is because of the nature of face expression changes. The objective is to achieve the highest possible level of precision. Patterns are found up to M = 3, which is selected by randomly selecting the depth of count in the resulting histogram using the 1 percent threshold. This is done in order to uncover patterns. There is a maximum of three patterns that might be found, which is denoted by the letter M.

When the deep learning method is utilised, the FC1 and FC2 Layers of both AlexNet and VGGFace produce feature vectors that have a total of 4096 dimensions. One of the prerequisites for recovering any of these is that each frame must first go through the necessary pre-processing operations before it can be sent out to the network. Here, the frames are modified to conform to the dimensions of 224 x 224 pixels for AlexNet and 227 x 227 pixels for VGGFace. In addition to this, it eliminates the typical image from the network. AlexNet makes use of pixels, but VGGFace makes use of colour channels when it comes to calculating averages.

Following the completion of the preprocessing of the frames, the subsequent phase consisted of transmitting the features that were generated by the filter responses through the networks in order to obtain them at the right concentrations. In order to determine which layer best exhibits the features required for depression analysis, separate evaluations of the two components will be conducted. We have established the FDHH threshold at one percent and M equal to five for the patterns in order to identify any variations that may exist. Following an evaluation of a sample that was chosen at random and subjected to FDHH, these parameters were chosen because they had a high pattern count of M = 10. We next proceed to analyse these ten patterns in order to discover the point at which they become indistinguishable from one another. It was discovered that the point in question was "01111110" in this particular instance. The conclusion that was reached as a result of this was that the appropriate pattern size is M= 5, which is the same as the most recent pattern finding.

Three separate data partitions are available for use in the Induced Emotion challenge, which makes use of the MediaEval16 database. It is through the use of the 9,800 samples that were submitted in that the algorithm is trained and tested. They are separated into categories, as indicated by the information that was provided by the organisers of the challenge. The challenge host will release 1,200 fresh samples around halfway through the course of the task. These samples will serve as the foundation for the testing partition that will ultimately be used. There is not a single one of these samples that has the original labelling. For the purpose of determining the effectiveness, we will make use of the Mean Squared Error as well as Pearson's Correlation Coefficient (PCC). By using Equation 6.7, one can obtain a method for determining the MSE that can be utilised. You will need to determine PCC in order to

$$pcc = \frac{cov(Y - \hat{Y})}{\sigma_y \times \sigma_{\hat{y}}} = \frac{\sum((\hat{Y} - \mu_{\hat{y}})(Y - \mu_Y))}{\sigma_{\hat{Y}} \times \sigma_Y}$$

Given a set of ground-truth labels Y and a set of predicted labels Y, we take Y and Y as the means of the labels, Y and Y as the standard deviations of the labels, and Y and Y as the standard deviations and means of the labels, respectively. This is done in order to determine the statistical significance of the labels. A reduction in the size of each image to 128 × 128 by 3 es is made by the framework in order to facilitate the auto-ability encoder's to train itself. It is done in this manner to guarantee that the sample sizes are consistent, that they do not surpass the memory capacity of the computer, and that they do not overlook essential spatial details due to their diminutive

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

size. In order to train an auto-encoder, it is necessary to make adjustments to the rk and ks hyperparameters of the network encoder.

a hundred different training practises;

In order to achieve the highest possible level of learning, we employ the stochastic gradient descent algorithm with the following parameters: batch size equals 25, learning rate equals 0.000001, momentum equals 0.9, and weight decay (regularisation) equals 0.00001. We begin by utilising Xavier's enhanced technique [89] to ensure that the network weights are chosen at random, and then we proceed to divide the data. Every lot is comprised of a single image taken from one of twenty-five video sequences that were chosen at random. In accordance with the guidelines of the competition, the training and testing portions should be distributed in an equal manner.

In the second step, a new sequence of features is created by giving the auto-encoder the frames from each visual sample that has been trained. On the other hand, the dimensions of each frame will continue to be the same, which are 128 by 128 by 3. There are five separate patterns that are produced when the FDHH algorithm is applied to each of the newly discovered feature sequences with a movement threshold of five percent. It may be concluded that the new FDHH feature size for each and every sample is either 128 by 128 by 3 by 5 or 128 by 128 by 15. After that, we will train two deep convolutional neural networks (DCNNs) using the FDHH features for each sample. One of the DCNNs will be used to forecast value, while the other will be used to forecast arousal. Initial configuration settings for both deep convolutional neural networks (DCNNs) are identical, including sizes, parameters, and hyperparameters. They are exactly the same as the encoders for the auto-settings, with the following significant differences:

As a result of the huge FDHH image, the batch size was decreased to ten, and the learning rate was modified to 0.0000001 seconds.

The loss function depicts the discrepancy between the two training procedures when stages 1 and 3 are compared to one another. You have the option of selecting between the MSE loss strategy or the Euclidean loss method throughout the setup process. The Mean Squared Error (MSE) loss is referred to as the DCNN MSE loss, while the Euclidean Distance (EC) loss is designated as the DCNN EUC loss. At this point in time, "audio" refers to anything that can be heard.

TYPES OF SUBSETS FOR ASSESSING DEPRESSION

Every frame of the example video is used to construct the EOH, LBP, and LPQ feature vectors. Three sets are constructed using these feature vectors. All of the data for these variables comes from the Depression Framework for Visual Data Phase 1. We begin by running the FDHH algorithm on the local feature vectors for EOH, LBP, and LPQ. This is done to ensure that all tests are run thoroughly. Finding out if each descriptor is compatible with the recovered FDHH feature requires a technique like this. The letters that make up its acronym are EOH FDHH, LBP FDHH, and LPQ FDHH. The final feature is the result of merging three separate local feature vectors. There will be 1,584 pieces overall when the feature is finished (384 plus 944 plus 256). The MIX FDHH feature is then generated by applying the FDHH feature to this feature vector. This component's design has been updated to accommodate M*1,584 separate parts. Those parts could end up in the Northwind movie as well as the Freeform one. The next step involves using a component vector with the dimensions 2 4,752= 9,510. To make this component vector, we combined the feature vectors for Northwind and Freeform.

The deep learning method retrieves the taught features from the AlexNet network by accessing the FC1 and FC2 layers. The FC1 and FC2 layers, which are levels 32 and 34, respectively, are achieved in VGGFace by means of extremely comparable methods. After being multiplied by 4096 (5 x 4096= 20,480), each of them is transformed into a vector with dimensions of 20,480 using the FDHH method. There will be a grand total of 4,960 possible dimensions after merging the Northwind and Freeform sets of capabilities. Their respective representations in AlexNet and VGGFace are A16 FDHH and A18 FDHH, and V32 FDHH and V34 FDHH, respectively.

The combined strategy benefits greatly from the extremely large feature sizes produced by the two methods when used in tandem. Results from a major component analysis indicate that 49 distinct subcomponents are now in play. This allows for a substantial diminution in the item's size. The full MIX FDHH matrix cannot be constructed with more primary components than this. The reason behind this is that there are a total of 9,550 feature components in the training set, which involves fifty samples. We can only use a certain amount of desktops, laptops, and servers for our investigation. Given this, the last aspect of the training set is fifty parts spread out over forty-nine classes. When looking at the Deep characteristics independently, a pattern does emerge. Then, we use the training set's covariance matrix to reduce the size of the development and testing sets to 49 pieces each.

Each set of features is evaluated and assigned a value between zero and one in order to get machine learning ready to be used. Regression becomes more effective and efficient with the introduction of feature value restrictions. Using all of the data in the training set allows us to focus on a certain feature. The process of computing the Rank normalisation can be seen in Equation 6.10. Xi is the normalised column in the feature matrix, min is the smallest value in the column, and max is the maximum value in the column. The column is denoted by Xi in this equation.

$$\hat{X}_i = \frac{X_i - \alpha_{min}}{\alpha_{max} - \alpha_{min}}$$

After we have completed the process of standardising the features, we will be able to examine the effectiveness of PLS and linear regression separately. Following that, the weighted sum rule is utilised in order to integrate the outputs of the regression analysis. By analysing the performance of the regression approach on the development partition, one is able to evaluate the effectiveness of the technique.

The approach for obtaining the desired mix of auditory characteristics from the qualities that are provided is referred to as the vocal branch method, which is illustrated in Figure. To begin, we will be utilising the section portions that are available in the shorter form. The total number of brief segments is 22,68, and the features that are given in Tables 6.1 and 6.2 are included in those segments. Prior to its release to the general public, this component, which has been given the name "Audio," will also be subjected to strict testing. There are many different audio properties that are utilised in order to generate unique sounds. Some of these attributes include

Flatness, Band1000, POV, PSY Sharpness, Shimmer, and ZCR. Out of the total of 2268 qualities, this accounts for 285 of them. The incidence in question is referred to as the (Comb). It is the responsibility of the MFCC to complete the audio features. To evaluate its efficacy in comparison to that of other solutions, there is no other way available. This is what is referred to as the (MFCC). The dimensionality of the characteristics is reduced through the use of principle component analysis (PCA), which is performed once the criteria have been set. By utilising the same parameters as the visual subtree, it is able to locate 49 personal computers, one of which is the decreased feature. The use of PLS and LR to the auditory data is followed by the formulation of predictions for depression scaling. As a result, decision-level fusion occurs.

In this article, we have compared the outcomes of various deep learning algorithms with regard to the recognition of facial expressions. F1 score, recall, accuracy, and precision are some of the evaluation measures that are included.

• The accuracy of a model is a measurement of how well it performs in terms of predicting outcomes in general.

• The accuracy of the model is evaluated based on how well it manages to recognise positive samples.

• A model's sensitivity or recall is a measurement of how well it can correctly identify all positive samples. This might be an important aspect of the model.

Given that it is the harmonic mean of recall and precision, the F1 score provides an accurate assessment of the effectiveness of the model under consideration.

Table 3: simulation analysis table for a project focused on Facial Expressions Recognition using deep learning algorithms:

Experiment No.	Deep Learning Algorithm Used	Dataset Size	Training Accuracy (%)	Validation Accuracy (%)	Test Accuracy (%)	Training Time (Minutes)	Remarks
1	Convolutional Neural Networks (CNN)	10,000	85	82	81	30	Satisfactory performance, but there is potential for improvement.
2	Recurrent Neural Networks (RNN)	10,000	80	78	76	40	Lower performance than CNN, longer training time.
3	Long Short Term Memory (LSTM)	10,000	82	80	79	45	Better than RNN but still less accurate than CNN.
4	Gated Recurrent Units (GRU)	10,000	83	81	79	42	Comparable to LSTM, slightly faster training.
5	CNN with Data Augmentation	10,000	88	86	85	35	Improved performance over basic CNN due to data augmentation.
6	Transfer Learning (VGG16)	10,000	91	88	87	25	High performance and efficient training time.
7	Ensemble Model (CNN + LSTM)	10,000	89	87	86	50	Good performance, but training time is higher.

It is important to note that the objective of this table is merely to provide an illustration. It is possible that the actual training times and accuracy rates would be determined by a number of elements, such as the model architecture, the complexity of the dataset, the available computer resources, and additional factors. The results analysis table includes a comparison of the performance of a number of different approaches. LSTM, Attention CNN, CNN, ResNet-50, and VGG-16 are subjected to evaluation in this section. By referring to the evaluation metrics that are presented in the table, you will be able to determine how well each strategy performed on the Facial Expressions Recognition test. It is recommended that you

replace the sample values currently included in this table with the actual assessment findings that you obtained from your research. It is possible that the measurements and approaches that are utilised will vary depending on the specific implementation and dataset that you have.

Table 4: results analysis table for Facial Expressions Recognition using a deep learning algorithm:

Method	Accuracy	Precision	Recall	F1 Score
CNN	0.85	0.84	0.86	0.85
ResNet-50	0.88	0.87	0.89	0.88
VGG-16	0.83	0.82	0.84	0.83
LSTM	0.79	0.77	0.81	0.79
Attention CNN	0.90	0.89	0.91	0.90
Proposed Method	Accuracy	Precision	Recall	F1 Score

Conclusion

"Facial Data Classification using Enhanced Local Binary Patterns (LBP) and Dynamic Range Local Binary Patterns (DRLBP) Algorithms" is an abstract that speaks about how far facial recognition technology has come. Facial data classification can be greatly improved by utilising the Enhanced LBP and DRLBP algorithms, which are extensively studied in this research. Improved Local Binary Patterns (Enhanced LBP) takes facial photos with more nuanced textural information and makes them more accurate than the original Local Binary Patterns (LBP). Applications such as security systems and human-computer interaction rely on precise and detailed representations of facial features, which this improvement enables. On the flip side, dynamic range LBP (DRLBP) adds a dynamic element to LBP. To improve the accuracy of face recognition, it adjusts to the variable pixel intensities seen in different areas of a face image. When faced with real-world obstacles, such as changing lighting conditions and facial expressions, this adaptive behaviour shines. By comparing these algorithms to more conventional approaches, the study proves that the former is more accurate, faster, and more flexible. Additionally, it delves into their potential uses in various settings, showcasing their compatibility with current face recognition systems. These state-of-the-art algorithms have the ability to utterly transform the face data classification industry, according to the study's consideration of future advancements and possible areas of use.

REFERENCES

[1]. K. Chen, T. Yi and Q. Lv, "LightQNet: Lightweight Deep Face Quality Assessment for Risk-Controlled Face Recognition," in IEEE Signal Processing Letters, vol. 28, pp. 1878-1882, 2021, doi: 10.1109/LSP.2021.3109781.

[2]. F. Mokhayeri, E. Granger and G. -A. Bilodeau, "Domain-Specific Face Synthesis for Video Face Recognition From a Single Sample Per Person," in IEEE Transactions on Information Forensics and Security, vol. 14, no. 3, pp. 757-772, March 2019, doi: 10.1109/TIFS.2018.2866295.

[3]. Z. Ma, Y. Liu, X. Liu, J. Ma and K. Ren, "Lightweight Privacy-Preserving Ensemble Classification for Face Recognition," in IEEE Internet of Things Journal, vol. 6, no. 3, pp. 5778-5790, June 2019, doi: 10.1109/JIOT.2019.2905555.

[4]. J. Muhammad, Y. Wang, C. Wang, K. Zhang and Z. Sun, "CASIA-Face-Africa: A Large-Scale African Face Image Database," in IEEE Transactions on Information Forensics and Security, vol. 16, pp. 3634-3646, 2021, doi: 10.1109/TIFS.2021.3080496.

[5]. S. Jia, X. Li, C. Hu, G. Guo and Z. Xu, "3D Face Anti-Spoofing With Factorized Bilinear Coding," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 31, no. 10, pp. 4031-4045, Oct. 2021, doi: 10.1109/TCSVT.2020.3044986.

[6]. X. Lv, C. Yu, H. Jin and K. Liu, "HQ2CL: A High-Quality Class Center Learning System for Deep Face Recognition," in IEEE Transactions on Image Processing, vol. 31, pp. 5359-5370, 2022, doi: 10.1109/TIP.2022.3195638.

[7]. Y. Wang and Q. Wu, "Research on Face Recognition Technology Based on PCA and SVM," 2022 7th International Conference on Big Data Analytics (ICBDA), 2022, pp. 248-252, doi: 10.1109/ICBDA55095.2022.9760320.

[8]. S. Shavetov and V. Sivtsov, "Access Control System Based on Face Recognition," 2020 7th International Conference on Control, Decision and Information Technologies (CoDIT), 2020, pp. 952-956, doi: 10.1109/CoDIT49905.2020.9263894.

[9]. Y. -X. He, "The influence of image enhancement algorithm on face recognition system," 2021 2nd International Conference on Computer Engineering and Intelligent Control (ICCEIC), 2021, pp. 20-24, doi: 10.1109/ICCEIC54227.2021.00012.

[10]. L. Fu and X. Shao, "Reseach and Implementation of Face Detection, Tracking and Recognition Based on Video," 2020 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS), 2020, pp. 914-917, doi: 10.1109/ICITBS49701.2020.00202.

[11]. B. Tej Chinimilli, A. T., A. Kotturi, V. Reddy Kaipu and J. Varma Mandapati, "Face Recognition based Attendance System using Haar Cascade and Local Binary Pattern Histogram Algorithm," 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184), 2020, pp. 701-704, doi: 10.1109/ICOEI48184.2020.9143046.

[12]. B. YILMAZER and S. SOLAK, "Cloud Computing Based Masked Face Recognition Application," 2020 Innovations in Intelligent Systems and Applications Conference (ASYU), 2020, pp. 1-5, doi: 10.1109/ASYU50717.2020.9259806.

[13]. Q. Zhai, X. Tian and Z. Feng, "Entrance Guard System Recognition Based on Face Block Feature Point Fusion,"

2021 2nd International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE), 2021, pp. 530-534, doi: 10.1109/ICBASE53849.2021.00104.

[14]. G. Singh and A. K. Goel, "Face Detection and Recognition System using Digital Image Processing," 2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), 2020, pp. 348-352, doi: 10.1109/ICIMIA48430.2020.9074838.

[15]. A. A. Mukhanbet, E. S. Nurakhov and T. S. Imankulov, "Hybrid Architecture of Face and Action Recognition Systems for Proctoring on a Graphic Processor," 2021 IEEE International Conference on Smart Information Systems and Technologies (SIST), 2021, pp. 1-4, doi: 10.1109/SIST50301.2021.9465971.

[16]. B. Gomathy, K. B. S. Sathya, S. J, S. S and S. K. S, "Face Recognition based student detail collection using OpenCV," 2022 8th International Conference on Smart Structures and Systems (ICSSS), 2022, pp. 1-4, doi: 10.1109/ICSSS54381.2022.9782211.

[17]. P. P D and P. Nath Singh, "Masked & Unmasked Face Recognition Using Support Vector Machine Classifier," 2021 IEEE International Conference on Mobile Networks and Wireless Communications (ICMNWC), 2021, pp. 1-4, doi: 10.1109/ICMNWC52512.2021.9688542.

[18]. H. Lin, H. Ma, W. Gong and C. Wang, "Non-frontal face recognition method with a side-face-correction generative adversarial networks," 2022 3rd International Conference on Computer Vision, Image and Deep Learning & International Conference on Computer Engineering and Applications (CVIDL & ICCEA), 2022, pp. 563-567, doi: 10.1109/CVIDLICCEA56201.2022.9825237.