Efficient image feature extraction using contrast information fractal dimension for hand sign recognition

Annalakshmi GANESAN[1]*, Gajalakshmi PALANISAMY[2]

[1] Department of Electronics and Communication Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Avadi, Chennai, Tamil Nadu, India

anumevlsi@gmail.com (*Corresponding author)

[2] Department of Computer Science and Engineering, CK College of Engineering and Technology, Cudallore, Tamil Nadu, India

Gajalakshmi.p@ckcet.edu.in

Abstract: Hand gesture recognition as the computer interpretation of human hand gestures used to assist people with disabilities. The texture-based feature descriptors of hand sign images do not have sufficient distinct features for recognition due to the change in geometric viewpoint and illumination invariant images. For instance, texture-based descriptors such as local binary patterns are sensitive to noise images. To overcome this problem, this paper proposes a novel global information extraction called contrast information fractal dimension (CIFD), which is based on contrast entropy and fractal dimension. The proposed CIFD method is used to obtain distinct features by analysing the edge quantity of hand sign images through measuring the information content, using non-linear filtering techniques and contrast entropy based on Weber and Devries-Rose contrast measurement. The extracted information from the hand edge images is applied to the convolutional neural network (CNN) for better hand sign recognition. The CNN is used to learn complex and non-linear relationships in static hand gesture images for recognition. The efficiency of the proposed system is verified by comparing various texture-based feature descriptors using standard databases. The hand sign recognition accuracy of the Jochen Triesch database for uniform and dark background environments are 99.50% and 95% and that of the National University of Singapore database (NUS-I) for black-white and colour background environments are 92.50% and 95% respectively.

Keywords: Convolution Neural Network, Contrast entropy information, Fractal dimension, Feature descriptor.

1. Introduction

A visual feature descriptor is a major cue for analysing the heterogeneous pattern of an image's surface through the properties of texture, shape and colour which provides sufficient descriptor for the fields of computer vision. Texture is the most familiar low-level descriptor in computer vision due to compact quantitative feature descriptors. However, many domains in computer vision such as image classification, content-based image retrieval and pattern recognition use texture descriptors which face major challenges due to geometric, illumination invariant and viewpoint changes (Nikan & Ahmadi, 2015). In the past, several approaches based on texture descriptors have been proposed to overcome these challenges and to achieve deterministic features in such domains. The fundamental principle of these approaches is to find the feature descriptors through the analysis of the texture surface features within a local window zone. The most popular conventional pattern descriptor is the local binary pattern (LBP) which has been applied for texture classification. The features of LBP are derived from thresholding, the pixel intensity of each neighbourhood pixel with the same intensity as the centre pixel, which are converted into a binary encoded form to act as additional details for image analysis. However, the LBP has discriminant features and suffers from noisy images under illumination conditions (Cusano, Napoletano & Schettini, 2014). It is enhanced by the local directional pattern (LDP) in which each pixel intensity from a local window texture pattern of an input image is assigned to a binary code of 8 bits (Tran et al., 2017). The LDP features are computed by the edge response with a Kirsch mask of eight directions in each pixel intensity position of an image. The binary pattern codes are then obtained by the intensity strength of magnitude from the local window. The complete modelling for the features of LBP (CLBP) is proposed (Guo, Zhang & Zhang, 2010) to overcome the issues of the LBP features of an image. The descriptor of CLBP encodes the features into a binary representation of the centre pixels and local pattern difference of magnitude transformation, which are used as

https://doi.org/10.33436/v36i1y202601

complementary components for enhancing the LBP descriptor. Ojansivu & Heikkila (2008) introduced local phase quantization (LPQ) to extract the local phase texture information in a local window. Ahmed & Hossain (2013) have presented a gradient local ternary pattern (GLTP) to describe the combination of the robust gradient for analysing the edge information by the Sobel operator and encoding the local pattern by using three levels of different coding techniques. Tan & Triggs (2010) have established a local ternary pattern (LTeP), which is the extension of LBP for handling the illumination conditions of an image for local features in binary code by three-level encoding. Here encoding is taking place by comparing the centre pixel with all other neighbouring pixels in a local window. Khan et al. (2013) proposed that the median ternary pattern (MTP) combine the median filter and gray scale image quantization of the local pattern into feature descriptors to achieve better accuracy in hand sign image recognition. In MTP, the local pixel pattern is used to determine the median and quantized into ternary code using specific threshold logic. The local pattern of grey scale integer pixel within threshold around the median is set as zero, pixels above to +1 and below to –1. The MTP quantized code in binary form is converted into a decimal pattern using the positive and negative median ternary pattern code. Finally, the histogram of feature descriptor is obtained by concatenating the positive and negative descriptors. Upadhyay, Lory & DeSouza (2025) proposed fixed a Scale Adaptive Robust Local Binary Pattern (SARLBP) to extract the optimal scale feature on all directions based on texture information. This effectively extracts the fine and macro structure information in detail compared with existing methodologies. Tekade et al. (2025) presented the Local Triangular Binary Pattern (LTBP) for person identification. In existing LBPs, the pairwise comparisons are made to form a unique feature information, proposing that LTBPs form triangular pixel groups within the local neighbourhood. Instead of only capturing center-to-neighbour disparities, each triangle collects second-order interactions among three pixels. Richer texture information is provided by this triangle encoding. Kumar et al. (2025) has proposed an orthogonal feature relationship which combines both LBP and bi-directional information (horizontal and vertical) to extract the feature value. This produces robust information against illumination and noise variance. The literature of recent techniques involves the idea of increasing the discriminative power for their applications through modifying and extending the local texture of an image, which serves as additional information for enhancing the local descriptors. In continuation of this, we proposed a new feature extraction method for hand sign recognition under various cases of different lighting conditions and geometric invariants. However, recent techniques of local pattern descriptors are good in various applications, the limitation of such techniques involving local descriptors resulting in less visual shape for the holistic texture of a hand sign image. Such a visual expression of holistic texture images is obtained by fractal dimension (FD) (Del-Pozo-Velázquez et al., 2025) measuring the texture image in terms of a homogeneous pattern which aggregates the micro textures of a hand sign image. Since the feature of a fractal dimension is inherent, the surfaces properties of micro texture are considered a highly reliable measure for deterministic feature extraction for hand sign recognition. In a FD, homogeneous hand sign images for feature extraction are enhanced through appropriate hand sign information by measuring the edge details of hand surface texture using a differential box counting (DBC) method. The functions of the DBC method are measuring the texture detail by counting the number of boxes in varying size in a non-linear kernel which operates on the image surface. Here, the entire texture surface is covered to obtain quality image details. Then, for strengthening the hand sign image details, we proposed a contrast information fractal dimension (CIFD) in which the final fractal image of hand signs is determined through contrast entropy for better recognition. In this paper, the contrast entropy implemented is derived from Weber and Devries-rose contrast for improving edge information of image under different illumination conditions.

The robust CIFD descriptors are applied to machine learning techniques for recognition of various hand signs. The different machine learning algorithms are linear support vector machines (SVM) (Berikol, Yildiz & Özcan, 2016), nonlinear kernel SVM (Zhang, Wang & Dong, 2014), fuzzy kernel based SVM (FKSVM) (Yang et al., 2010), extreme machine learning (ELM) (Chen et al., 2022) and deep neural networks (Kothadiya et al., 2022). The practical issues in conventional classifiers are complexity, memory requirements, dimensionality curse, generalization performance, classification time, training time and error. To overcome such drawbacks, Huang (2014) proposed a minimum training error by solution of minimum norm least square and thereby achieving good generalization performance. Samat et al. (2014) proposed a solution for high dimensional based extreme learning machine (ELM) in which ensemble classifiers such as Bootstrap aggregation and AdaBoost are integrated with ELM to handle hyperspectral data. Hernandez-Hernandez & Rubio-Solis (2025) proposed dimension reduction in the framework of linear and non-linear such as ELM auto-encoder and sparse ELM auto-encoder to achieve less training time and discriminative capability. In this proposed work, the CNN model is used to detect the pattern and analyse the images for hand sign recognition. The new feature descriptor called contrast information fractal dimension (CIFD) combines the fractal dimension with the contrast entropy based on the Weber and Devries-Rose method. The fractal dimension method is used to enhance the edge details of the images by measuring the roughness parameter of the corresponding images through box counting techniques.

The contribution of the work is summarized as follows:

1. The contrast information fractal dimension (CIFD) based feature descriptor is proposed to enhance the edge details of the hand sign.
2. In an illumination condition, the magnitudes of edge detail images are not clearly shown. Hence, fractal dimension is applied to enhance the edge quality details through filtering techniques for achieving discriminant feature representation.
3. The experiment is conducted on various hand sign databases such as Jochen Triesch and NUS-I. The proposed technique performs better than the other conventional feature descriptor.

Additionally, part of this work is constructed as follows: Section 2 describes the details about the proposed methodology and Section 3 details the hand sign recognition using CNN models. The experimental results of the proposed method are explained in Section 4, followed by the conclusion in Section 5.

2. Feature extraction

The features are analysed under an illumination invariant of hand sign images for recognition which needs an effective tool to enhance the edge details of an image. In the illumination condition, the magnitudes of edge detail images are not clearly shown. Hence, the fractal dimension is applied to enhance the edge quality details through filtering techniques for achieving discriminant feature representation.

In order to transfer the hand sign images $\left(IM\right)$ of size $A \times B$ to the fractal image (FD), first the 3D matrix is computed to represent the number of boxes required to cover the entire image surface as follows:

$$Matrix_{\mathrm{dim}} = (x, y, \mathrm{dim}) \tag{1}$$

where (x, y) represents the pixel coordinates and denotes the scaling level. The 3D matrix is generated using the weighted local summation of the window size $A \times B$ with the center point (x, y) overlaying the image $\left(IM\right)$ at each pixel (x, y) as defined in Equation (2):

$$Matrix_{\mathrm{dim}}(x, y, \mathrm{dim}) = \sum_{m=-k}^{k} \sum_{n=-l}^{l} \ker(m, n) IM_{A \times B}\left(x + m, y + n\right) \left[\frac{S_{\max}}{S}\right]^2 \tag{2}$$

where dim is the matrix dimension, S is the scaling factor along with greatest value $S_{\max}$ that indicates the quantity of the specific structure pattern around its texture surface for hand images. This work proposes the minimum and median filter as nonlinear kernel filter factors $\ker(m, n)$ to enhance the edge information and preserve edge quality of hand images. These two filters are combined to produce an efficient gradient of an image for analysing the edge details which is given in Equation (3):

$$\ker(m,n) = round\left[\frac{\rho_{median} - \rho_{\min}}{S}\right] + 1 \tag{3}$$

where $\ker(m,n)$ is a non-linear kernel with varying size of $S \times S$ and $k, -k, l, and -l$ are positive integers to compute kernel on each image pixel. The function ρ_{median} is a median filter which is determined by sorting the pixel values in the kernel window, and then it is replaced with the pixel of middle value. The function $\rho_{\min}$ is the least intensity of neighbouring pixel values in the kernel block. Moreover, to obtain good quantity details of hand images, it is necessary to adjust the visibility level between the foreground and background by the contrast level of $Matrix_{\dim}$ images. In this paper, the factor of contrast property from the Weber and Devries-Rose contrast is derived for well-defined quantity edges. The level of visibility information is mainly based on the spatial distribution of light intensity of an image. Absolute brightness is related to the amount of light stimulus. The relative brightness of an image is determined by the difference between luminance of the maximum and minimum intensity, which is measured by contrast. The contrast of an image is either the object brightness value becoming bright or dark based on background or surrounding effects using the threshold intensity. Depending on the characteristics of threshold and background intensities, the relationships between such intensities are derived in the Devries-Rose method:

$$Log\left(\Delta T_1\right) = \log\left(C_1\right) + \left(\frac{\log\left(B_I\right)}{2}\right) \tag{4}$$

where ΔT_1 , B_I and C_{rl} are the threshold intensity, background intensity and Weber contrast. The entropy of the contrast relationship C_{rl} is then determined by Equation (5):

$$C_{rl} = \log\left(\Delta T_I\right) \tag{5}$$

$$E_c = \frac{1}{row \times col} \sum_{i=1}^{row} \sum_{j=1}^{col} alpha\left(C_{rl}\right)_{i,j}^{alpha} \times \ln\left(C_{rl}\right)_{i,j} \tag{6}$$

where E_c is the contrast entropy for each location block $\left(C_{rl}\right)_{i,j}$, scaled through $alpha$ which is constant for the entire image. To achieve a robust illumination invariant, the contrast information entropy of an image E_c and $Matrix_{\dim}$ is combined for novel descriptors. The contrast information entropy of an image E_c is calculated by using Equation (6) from the matrix to achieve unique feature values.

The fractal slope of linear regression is generated by the matrix to represent the fractal dimension $F_{\dim}\left(x, y\right)$ with a scale factor S . The image Z consists of $Z_1, Z_2, Z_3 .....Z_{A \times B}$ in which each $i = 1, 2, 3.....A \times B$ is a vector size $\dim = S_{\max} - 1$. The $F_{\dim}\left(x, y\right)$ is generated by the fractal slope of the linear regression and this slope is computed by the sum of squares as follows.

$$F_{\dim}\left(x, y\right) = \frac{N_z\left(x.y\right)}{N_0} \tag{7}$$

$$Nz\left(x, y\right) = \sum_{S=1}^{S_{\max}} \log(s)\left(Z\left(x.y\right), S\right) - \frac{\sum_{S=1}^{S_{\max}-1} \log(S)\left(Z\left(x.y\right), S\right) \sum_{S=1}^{S_{\max}} \log\left(S\right)}{S_{\max} - 1} \tag{8}$$

$$N_0 = \frac{\left(\sum_{S=1}^{S_{\max}-1} \log(S)\right)}{S_{\max}-1} \tag{9}$$

The log-based transformation is used to brighten the image and lessen the lighting impact. The values of bright pixels are compressed while the values of dark pixels are expanded using this log function. Then combine the $F_{\dim}$ and contrast information entropy of an image E_c to provide enhanced edge details in images for recognition of hand gestures. The proposed technique procedure is presented in Algorithm 1.

Algorithm 1. Extract CIFD features from hand sign gesture image

Input: Hand sign image $IM_{A\times B}$.

Output: Contrast Information Fractal Dimension (CIFD) Image.

1. Declare , $S_{\min} = 2$, $S_{\max} = 11$ and $\dim = S_{\max} - 1$
2. For $S = S_{\min} : S_{\max}$
3. Determine the $Matrix_{\dim} = (x, y, \dim)$ of FD value

$$Matrix_{\dim}(x, y, \dim) = \sum_{m=-k}^{k} \sum_{n=-l}^{l} \ker(m,n) IM_{A\times B}\left(x+m, y+n\right)\left[\frac{S_{\max}}{S}\right]$$

4. Apply nonlinear kernel filter factors $\ker(m,n)$ to enhance the edge information

$$\ker(m,n) = round\left[\frac{\rho_{median} - \rho_{\min}}{S}\right] + 1$$

5. Compute the fractal dimension image using

$$F_{\dim}\left(x, y\right) = \frac{N_z\left(x.y\right)}{N_0}$$

6. Compute contrast entropy E_c by Equation (6) from the $Matrix_{\dim}(x, y)$ and combine with $F_{\dim}(x, y)$
7. End

The fractal dimension (FD) effectively captures the overall hand contour and boundary irregularities. Global structural complexity and boundary abnormalities are clearly visible in this feature descriptor. This technique is insensitive to small edge features and local texture variations, which are essential for correctly differentiating subtle hand gestures. Contrast entropy clearly shows the gray scale distribution of the pixels. It effectively captures the fine texture details of the fingers. However, spatial geometry and global hand structure, both necessary for differentiating between hand motions with comparable local textures but dissimilar spatial arrangements, cannot be efficiently encoded by contrast entropy alone. When the Fractal Dimension (FD), nonlinear filtering, and Contrast Entropy (CE) are combined to create a powerful and discriminative feature representation by capturing local texture details, global hand structure and enhanced edges. This ensures robust multi-scale encoding, precise edge discrimination, stability under noise and illumination changes, improved separation between gesture classes and reduced variability within classes, all of which contribute to accurate and dependable recognition. The proposed CIFD enriches the edge information for robust feature descriptors by the effectiveness of the FD through contrast entropy of images. The computation of CIFD is illustrated in Figure 1. The proposed technique output for the various dataset is presented in Figure 2. It is observed that the proposed technique shows the edge fractal features and contrast entropy working together to capture both local variations intensity and global structural complexity, resulting in a powerful and complementary visual representation. Contrast entropy accurately encodes fine-grained local contrast, edge sharpness and micro-textural characteristics, whereas fractal features efficiently characterize the uniform, complex and scale-invariant patterns of objects, offering robust characterisation of shape and texture. The integrated images are then fed into the CNN classifier for hand sign recognition.

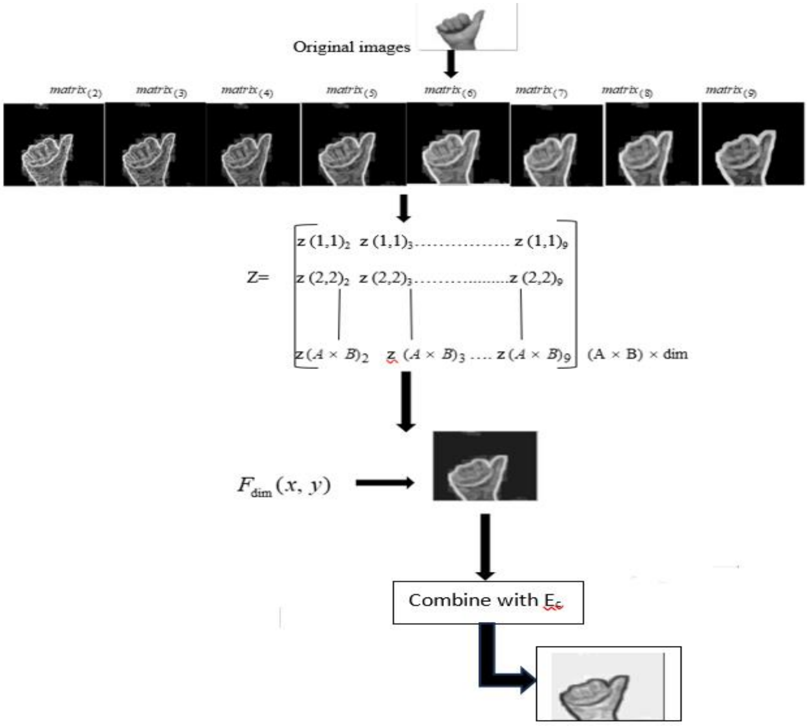


Figure 1. Procedural steps of the proposed feature descriptor

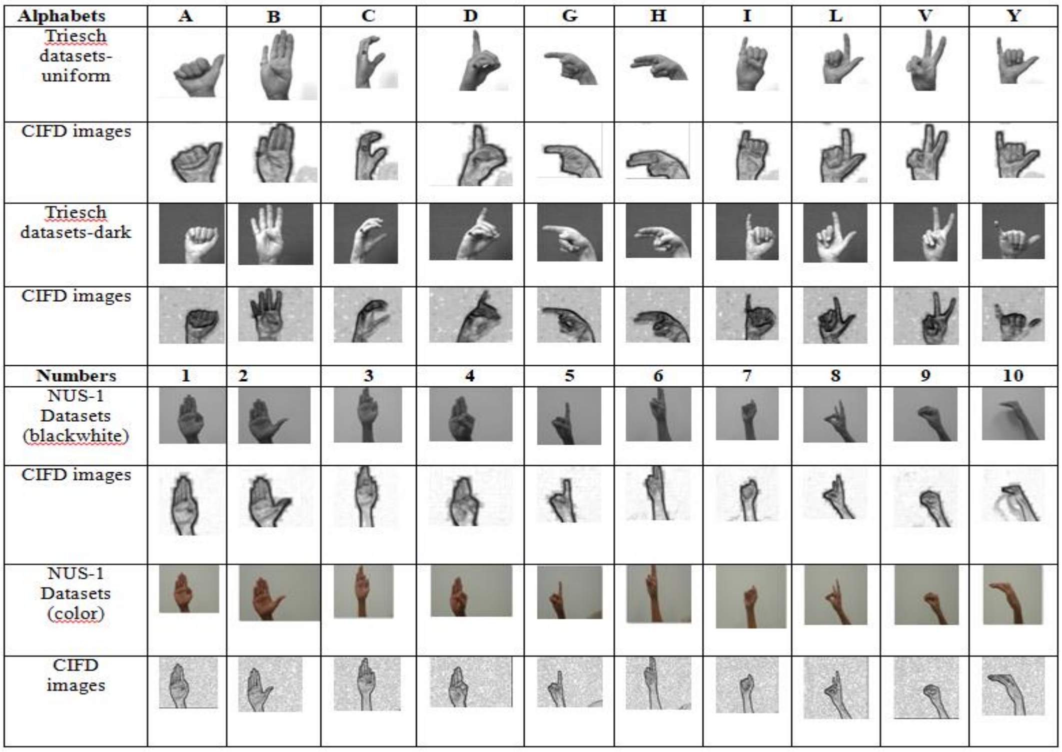


Figure 2. CIFD images from Triesch datasets (uniform, dark) and NUS-1 datasets (black-and-white and colour backgrounds)

3. Hand gesture recognition

Finally, the extracted images are fed into the Convolution Neural Network (CNN) for the final classification. The learning processes in this system are inspired by the way neurons process information and respond to external stimuli, acquiring information with the support of sensory organs. The deep learning model predicts the output by processing the input of raw images through hidden layers of non-linear transformations. The function behind the deep learning system of neural network describes each neuron input of each hidden layer, which is determined from the layer of previous values. Such a process is dependent on the parameters of weights and bias values with activation function for non-linear transformation performance. When this process occurs in input and each hidden layer, the information in this layer becomes more complex and thus the deep learning systems use several non-linear transformation layers to solve complex information rather than using simple neural networks.

3.1. Convolutional Neural Network (CNN) architecture

A convolutional Neural Network (CNN) is one part of a deep learning system which is used mainly in pattern recognition (Hu & Wang, 2020). The CNN consists of three layers as input, several hidden and an output layer. Typically, the hidden layers in the CNN model have convolutional layers, pooling layers and final layers. In the CNN model, the input of each image is processed through a sequence of convolutional layers with kernel (filters) and pooling in between the layers and then processed to the final layer. The image object is classified by computing probability values between 0 and 1 using the Softmax function. In this CNN, the output hand sign image of local descriptors is given as input image to the CNN model of the deep learning classifier. The input to each layer in the CNN is represented as $p \times q \times r$, where p is width, q is height and r is depth (i.e., the number of channels), which is three for an RGB image. Each of the convolutional layers for the various kernels (filters) are used and mentioned as f_k. The dimension of f_k is described as $a \times b \times s$, which is similar but smaller than the input of the local descriptor images. Feature maps are generated in the convolutional layers by convolving the kernel basis of the local window filter f_k with input image and followed by the ReLU (rectified linear unit) non-linear activation functions. The feature maps are determined by the following expression:

$$feature\_map_{p,q,f_k} = \max\left(W_{f_k}^T A_{p,q}, 0\right) \tag{10}$$

where A is the input image at pixel location and f_k is the index in the channel of the feature map. Down-sampling in every feature map of the sub-sampling layers leads to a reduction of the parameter in the network model which is tuning and training the data, thereby handling the issues of under-fitting and over-fitting data. Such down-sampling is applied by a pooling function of either max or average pooling used. Finally, the average of each feature map in the previous layer is computed to obtain the resulting feature vector, which is then passed through the softmax function in output layer to classify the hand sign categories (Hou & Wang, 2019). Through global average pooling, feature maps are easily interpreted and closer to native to the corresponding hand sign image categories. The architecture diagram of the CNN is shown in Figure 3.

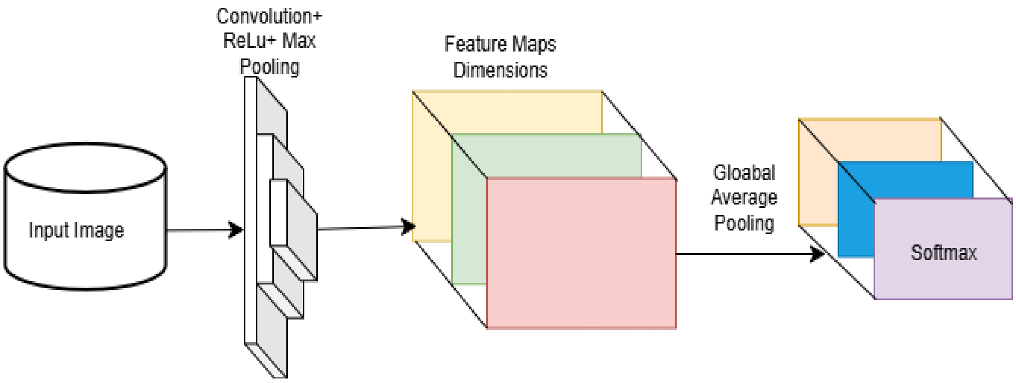


Figure 3. CNN architecture

4. Results and discussion

In this section the performance of the proposed method is analysed and validated with other state-of-the-art algorithms. The proposed method performances are evaluated on the publicly available datasets Jochen Triesch under uniform and dark background (Triesch, Malsburg & Marcel 2014), National University of Singapore (NUS-1) datasets under black-white and color background NUS1 (2025). The following subsections describe the dataset and evaluation of the proposed method performance.

4.1. Dataset details

The dataset of hand sign alphabets from Jochen Triesch, comprising light and dark conditions with illumination and geometric invariants, was used for recognition. This dataset has 10 different postures, each class containing data from 24 unique individuals with different conditions such as complex, uniform dark and uniform light backgrounds. The uniform dark and light datasets are considered for evaluation. The National University of Singapore (NUS-1) dataset includes 10 different postures with 24 sample images in each class. The images are obtained by altering the size and position of the hand with the frame. The hand postures are captured with less variation in each class to increase the difficulty of the recognition task. The dataset details are shown in Table 1.

Table 1. Details of standard hand sign datasets

Datasets	Hand sign images category	Number of sign images	Total sign images	Size of image
Jochen Triesch under Light illumination	10	24	240	128×128
Jochen Triesch under Dark illumination	10	24	240	128×128
NUS-1 (black-white)	10	24	240	160×120
NUS-1 (color)	10	24	240	160×120

4.2. Evaluation of the proposed method’s performance

The various pre-processed local descriptors results, such as LBP, LDP, complete CLBP, LPQ, GLTP, LTeP and MTP, are shown in Figure 4, where the output of various descriptors is observed through the indistinct foreground image of the hand sign alphabet representing ‘A’ due to noisy imaging under illumination invariants from the Jochen Triesch light and dark backgrounds as well as the NUS-1 dataset of black and white and color backgrounds. However, the CLBP and

MTP descriptors show a distinct output of hand signs due to using the principle of the mean filter, resulting in the same effects in LTeP descriptors. The hand signs through CLBP, MTP and LTeP descriptors achieve less performance in recognition of signs due to a lack of holistic data information. It is observed that enough deterministic features of hand signs are needed to improve the holistic texture details of hand signs for good recognition. In this work, such deterministic features are achieved through distinct hand signs of CIFD descriptors.

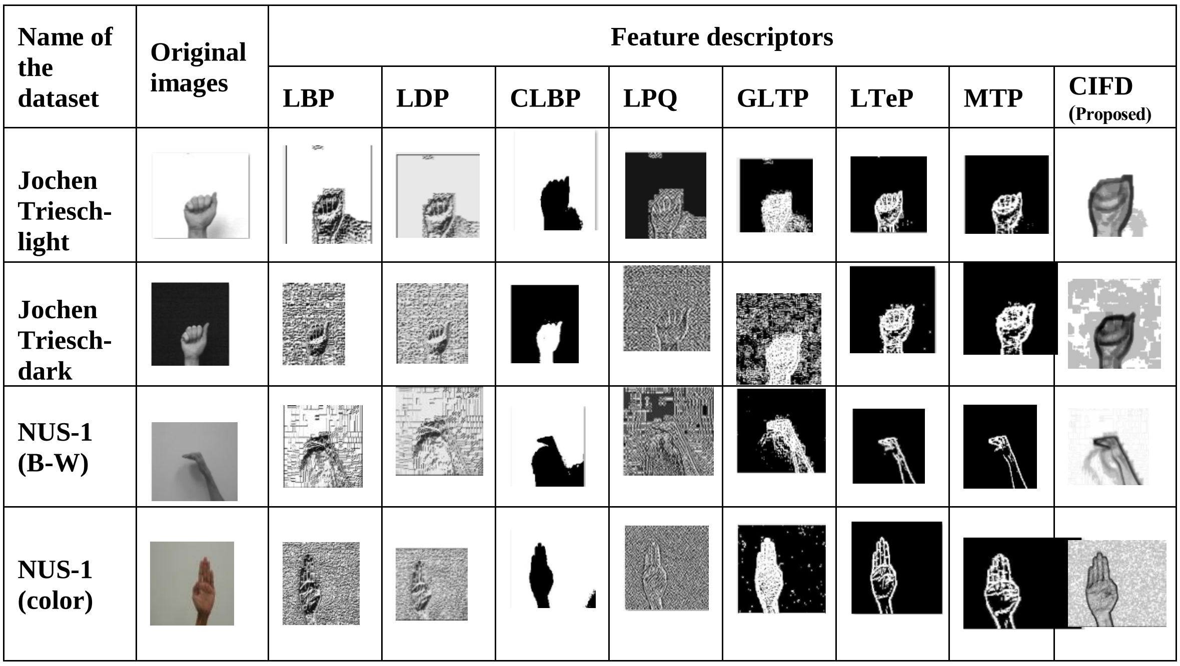


Figure 4. Outputs of various local descriptors for the "A" sign from the Triesch dataset (light and dark backgrounds) and the "10" and "1" signs from the NUS-1 datasets (black-and-white and color backgrounds)

The various state-of-the-art techniques for local and CIFD descriptors' output of hand sign images are given to the deep learning classifier. Here, the CNN model learns and preserves the positional characteristics of the local descriptor features. Moreover, this model can learn the characteristics of hand sign texture features better than the pixel values level in an original image. In this paper, throughout the classification of hand sign images, the experimental setup considers 80% of the images for training and the remaining images for testing and validation. For the optimization of the CNN model, hyper parameters are used, as shown in Table 2. The parameters are kept same throughout the simulation.

Table 2. The hyperparameters used in the CNN model

Batch size	32
Number of Epochs	110
Model Optimizer	SGD
Number of layers in the Model	154
Loss Function	Categorical Cross entropy
Activation function in between layers	ReLU
Activation function in output layer	Softmax
Preprocessing	Local descriptor using CIFD

The CNN model takes 32 samples for each batch to train the network, repeating the process until all the samples are trained in an iteration. For improved performance and loss minimization, the learning rate and weight properties are updated by the Stochastic Gradient Descent (SGD). The number of epochs evaluates the number of cycles that the algorithm performs on the training data. The loss is finding the error in the neural network, which is predicted by the loss function using the categorical cross entropy. The final parameter of activation functions is used to perform complicated processes on the hidden layers to obtain non-linear features in the neural network of the output layer using ReLU and SoftMax (Gupta & Porwal, 2016).

The classification's performance is evaluated using a five-fold cross-validation. To increase the size of the data augmentation, the original dataset is replicated 10, 20 and 30 times prior. Afterwards, the augmentation techniques are applied to each copy of the dataset. Augmentation techniques such as scaling, shearing, translation and flipping are used to improve the more variable training dataset. The comparisons of various hand-crafted features such as LBP, CLBP, LDP, LPQ, GLTP, LTeP, MTP and proposed CIFD are shown in Table 3. It is observed throughout the table that the LBP achieves the lowest classification rate of 45 % on the Jochen Triesch dark dataset, where the proposed CIFD feature descriptor achieves the 95% on that particular dataset. The proposed CIFD feature descriptor achieves the highest classification rate of 99.50 on the Jochen Triesch uniform dataset.

Table 3. Comparison of various feature descriptors

Feature Descriptor	Jochen Triesch Uniform	Jochen Triesch Dark	NUS-I Black and White	NUS-I Color
LBP	78.00	45.00	50.00	77.50
LDP	85.00	50.00	48.00	57.50
CLBP	82.00	90.00	85.50	89.00
LPQ	78.50	65.00	50.00	50.00
GLTP	65.00	68.00	60.00	73.00
LTeP	85.00	82.00	82.00	87.50
MTP	85.37	88.37	79.00	94.00
proposed CIFD	99.50	95.00	92.50	95.00

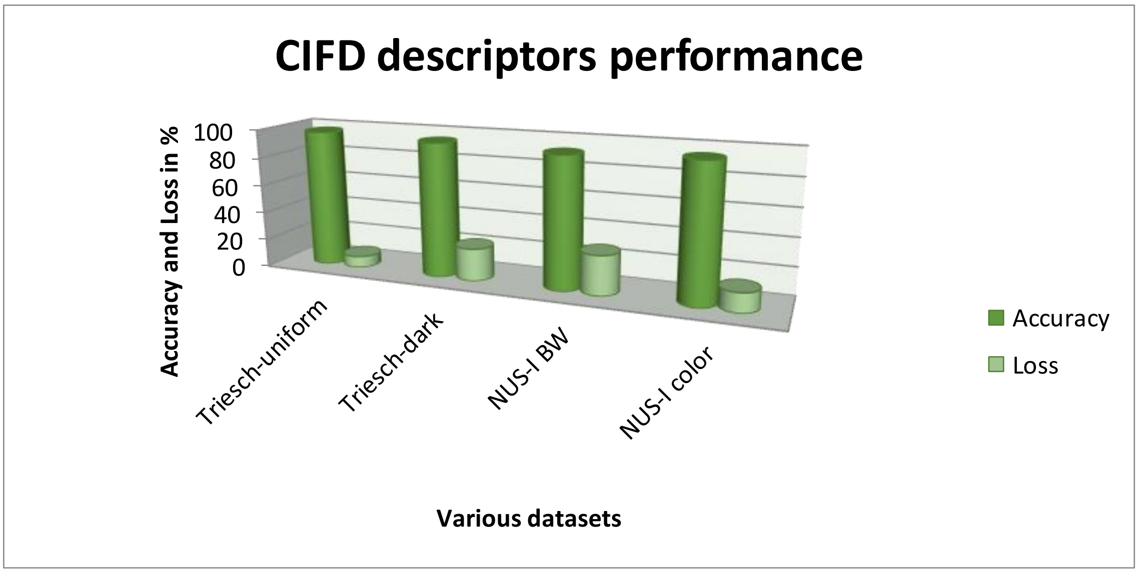
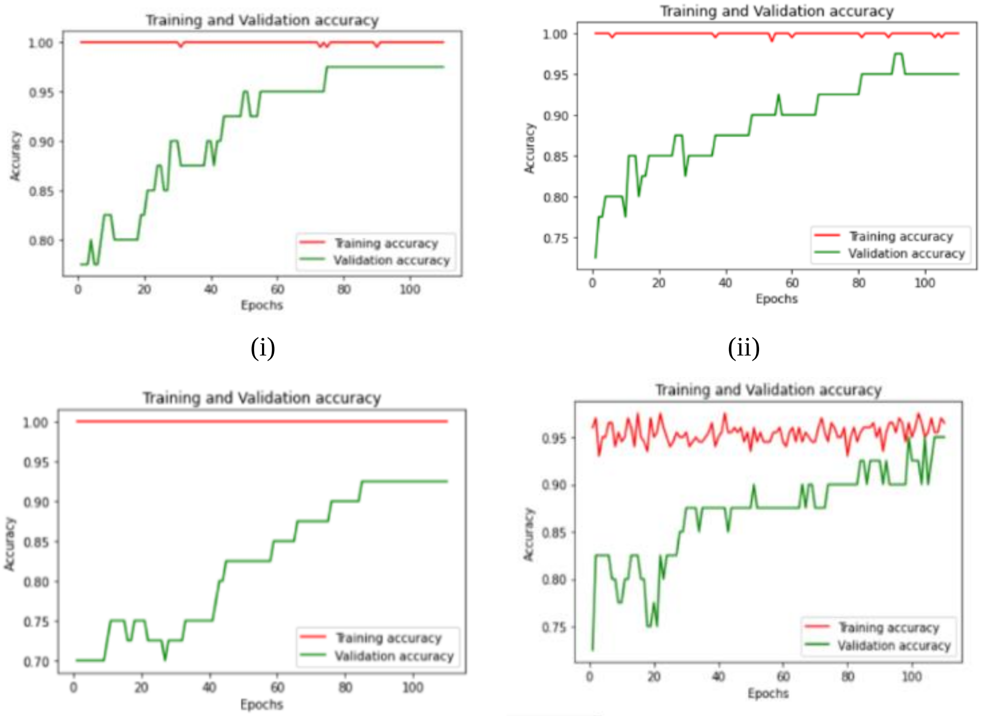


Figure 5. Performance comparison of CIFD descriptors on various datasets

As shown in Figure 5, the CIFD based CNN classifier outperforms the various local descriptors and achieves high accuracy in training and validation of hand sign data. Moreover, above 30 epochs, the CIFD feature descriptors gradually increase towards high recognition accuracy of hand sign images. This is due to the measuring of texture complexity such as roughness and smoothness level for images using DBC through a non-linear filter, as well as enriching the edge details quantity of images using a contrast property for distinguishing features. The metric of validation loss of the CIFD is almost zero when recognition accuracy increases. The validation and training accuracy of the proposed technique is shown in Figure 6.



(i) (ii)

(iii) (iv)

Figure 6. Recognition accuracy based on the CIFD descriptors from various datasets
(i) Triesch_uniform (ii) Triesch_dark (iii) NUS-1 under black and white (iv) NUS-1 under color

It is shown in Table 4 that when the unprocessed image is fed to the CNN classifier, it reaches a peak classification accuracy of 75% on the NUS-I black and white dataset. The Sobel filter is then applied to the original images, improving the edges of objects and structural details of the hand. The prepared images are subsequently provided to the CNN classifier and achieve a classification rate of 78.50% on the NUS-1 (black and white dataset). To further analyse the effectiveness of the proposed algorithm, the Retinex algorithm is applied in the preprocessing stage, which helps normalizing the illumination of the input image by separating the reflectance of that particular image. This enhances the contrast and reduces the shadows present in the image and achieves an 80.60% accuracy on the particular dataset.

Table 4. Comparison of CNN performance with different algorithms

Dataset	**CNN**	**CNN with Sobel Filter**	**CNN with Retinex algorithm**	**CNN with the proposed CIFD**
Jochen Triesch uniform	72.36	74.65	78.45	99.50
Jochen Triesch dark	65.77	75.43	74.00	95.00
NUS-I black and white	75.00	78.50	80.60	92.50
NUS-I color image	72.00	75.00	73.60	95.00

Here, the factors used in CIFD descriptors fulfil the quantized information of images to seek deterministic features of hand sign images. Such factors used in CIFD gives distinguish features which insist the parameters model of CNN to easily learn the characteristics of hand sign images better than the pixel level of an image. Moreover, the structure of CNN model along with various optimization functions such as epochs, batch size, pooling layer, model optimizer, loss and activation functions are used which lead to recognize the hand sign in better manner. The CIFD descriptors-based CNN model achieves maximum accuracy of testing samples of hand sign recognition from various datasets of Triesch-uniform, Triesch-dark, NUS-1 (Black-white, color background).

5. Conclusion

The descriptors of various local patterns of hand images are calculated by capturing the structured information in the form of two factors, such as encoding and quantization in the local window of an image, which are determined by the differences of the neighbour pixels. The local feature descriptors rrestrict the effective extraction of magnitude-related features in hand sign images. This leads to the average power of discriminant features for hand gesture recognition. In order to strengthen the hand sign pattern, the detailed information of the hand sign region is obtained, producing high discriminative features for robust hand sign recognition under the illumination change. Thus, a new CFID local descriptor is proposed for quantifying the homogeneous texture through fractal measurement as well as focusing the edge details of hand sign images through contrast property for well-known feature descriptors. These feature descriptors of hand sign images under various background conditions are applied to the CNN model for hand sign recognition. Therefore, the recognition of hand sign images through CIFD pre-processing with the CNN classifier achieves high recognition accuracy compared to the existing models. In the future it will be beneficial to employ optimization techniques in the classifier, further improving the recognition rate.

Author contributions

Conceptualization: A.G. and G.P.; Data Curation: A.G. and G.P.; Supervision: A.G.; Validation: A.G and G.P.; Writing—original draft: A.G. and G.P.; Writing—review and editing: A.G. and G.P. All authors have read and agreed to the published version of the manuscript.

Submission received: 22 July 2025; Revised:28 January 2026; Accepted: 04 February 2026; Published: 31 March 2026.

REFERENCES

Ahmed, F. & Hossain, E. (2013) Automated Facial Expression Recognition Using Gradient - Based Ternary Texture Patterns. *Chinese Journal of Engineering*. 2013(1), 831747. https://doi.org/10.1155/2013/831747.

Berikol, G. B., Yildiz, O. & Özcan, İ.T. (2016) Diagnosis of Acute Coronary Syndrome with a Support Vector Machine. *Journal of Medical Systems*. 40(4), 84. https://doi.org/10.1007/s10916-016-0432-6.

Chen, Y., Zhang, C., Liu, C. et al. (2022) Atrial fibrillation detection using a feedforward neural network. *Journal of Medical and Biological Engineering*. 42(6), 63-73. https://doi.org/10.1007/s40846-022-00681-z.

Cusano, C., Napoletano, P. & Schettini, R. (2014) Combining local binary patterns and local color contrast for texture classification under varying illumination. *Journal of the Optical Society of America A*. 31(7), 1453-1461. https://doi.org/10.1364/JOSAA.31.001453.

Del-Pozo-Velázquez, J., Aguiar-Pérez, J. M., Chamorro-Posada, P., Pérez-Juárez, M. Á., Wang, X. & Casaseca-de-la-Higuera, P. (2025) Smoke detection in images through fractal dimension-based binary classification. *Digital Signal Processing*. 166, 105346. https://doi.org/10.1016/j.dsp.2025.105346.

Guo, Z., Zhang, L. & Zhang, D. (2010) A completed modeling of local binary pattern operator for texture classification. *IEEE transactions on image processing*. 19(6), 1657-1663. https://doi.org/10.1109/TIP.2010.2044957.

Gupta, S. & Porwal, R. (2016) Appropriate contrast enhancement measures for brain and breast cancer images. *International journal of biomedical imaging* 2016(1), 4710842. https://doi.org/10.1155/2016/4710842.

Hernandez-Hernandez, R. A. & Rubio-Solis, A. (2025) A Hybrid Multilayer Extreme Learning Machine for Image Classification with an Application to Quadcopters. To be published in *Computer Vision and Pattern Recognition.* [Preprint] https://doi.org/10.48550/arXiv.2507.08047. [Accessed: 15th February 2025].

Hou, S. & Wang, Z. (2019) Weighted Channel Dropout for Regularization of Deep Convolutional Neural Network. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 33(01), 8425-8432. https://doi.org/10.1609/aaai.v33i01.33018425.

Hu, B. & Wang, J. (2020) Deep learning-based hand gesture recognition and UAV flight controls. *International Journal of Automation and Computing.* 17(1), 17-29. https://doi.org/10.1007/s11633-019-1194-7.

Huang, G. B. (2014) An insight into extreme learning machines: random neurons, random features and kernels. *Cognitive computation*. 6(3), 376-390. https://doi.org/10.1007/s12559-014-9255-2.

Khan, A., Bashar, F., Ahmed, F. & Kabir, M. H. (2013) Median ternary pattern (MTP) for face recognition. In *2013 International Conference on Informatics, Electronics and Vision (ICIEV), 17-18 May 2013, Dhaka, Bangladesh.* IEEE. pp. 1-5. https://doi.org/10.1109/ICIEV.2013.6572546.

Kothadiya, D., Bhatt, C., Sapariya, K. et al. (2022) Deepsign: Sign language detection and recognition using deep learning. *Electronics*. 11(11), p1780. https://doi.org/10.3390/electronics11111780.

Kumar, K. N., Nirmal, G., Vaghulade, M., Patil, S. A. & Verma, A. (2025) Ortho-LBP: novel features for enhanced pattern recognition through bitwise logical fusion of bi-directional orthogonal information with LBP. *International Journal of Information Technology*. 1-12. https://doi.org/10.1007/s41870-025-02747-y.

National University of Singapore (2025) NUS https://scholarbank.nus.edu.sg/entities/dataset/af106bc7-7be6-44b8-937d-6dc338b92173. [Accessed: 15th February 2025].

Nikan, S. & Ahmadi, M. (2015) Local gradient-based illumination invariant face recognition using local phase quantisation and multi-resolution local binary pattern fusion. *IET image processing*. 9(1), 12-21. https://doi.org/10.1049/iet-ipr.2013.0792.

Ojansivu, V. & Heikkilä, J. (2008) Blur insensitive texture classification using local phase quantization. In *International conference on image and signal processing*. Berlin, Heidelberg: Springer Berlin Heidelberg. (pp. 236-243). https://doi.org/10.1007/978-3-540-69905-7_27.

Samat, A., Du, P., Liu, S. et al. (2014) E^2LMs: Ensemble Extreme Learning Machines for Hyperspectral Image Classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. 7(4), 1060-1069. https://doi.org/10.1109/JSTARS.2014.2301775.

Tan, X. & Triggs, B. (2010) Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE transactions on image processing*. 19(6), 1635-1650. https://doi.org/10.1109/TIP.2010.2042645.

Tekade, A., Vijayan, T., Karthik, B. & Mahajan, A. (2025) Person identification using novel local triangular binary pattern-based texture descriptor. *EURASIP Journal on Advances in Signal Processing*. 2025(1), 7. https://doi.org/10.1186/s13634-025-01213-y.

Tran, C. K., Tseng, C. D., Chao, P. J. et al. (2017) Face Recognition under Varying Lighting Conditions: A Combination of Weber-face and Local Directional Pattern for Feature Extraction and Support Vector Machines for Classification. *J. Inf. Hiding Multim. Signal Process.* 8(5), 1010-1020.

Triesch, J., Malsburg, C. & Marcel, S. (2014) *Hand posture and gesture datasets: JochenTriesch static hand posture database*. https://www.idiap.ch/webarchives/sites/www.idiap.ch/resource/gestures/ [Accessed: 15th February 2025].

Upadhyay, P. C., Lory, J. A. & DeSouza, G. N. (2025) SARLBP: scale adaptive robust local binary patterns for texture representation. *IEEE Transactions on Image Processing*. 34, 969-981. https://doi.org/10.1109/TIP.2025.3529376.

Yang, X., Zhang, G., Lu, J. & Ma, J. (2010) A kernel fuzzy c-means clustering-based fuzzy support vector machine algorithm for classification problems with outliers or noises. *IEEE Transactions on fuzzy systems*. 19(1), 105-115. https://doi.org/10.1109/TFUZZ.2010.2087382.

Zhang, Y. D., Wang, S. & Dong, Z. (2014) Classification of Alzheimer disease based on structural magnetic resonance imaging by kernel support vector machine decision tree. *Progress in Electromagnetics Research*. 144, 171-184. https://doi.org/10.2528/PIER13121310.

Annalakshmi GANESAN is an Assistant Professor in the Department of Electronics and Communication Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology. She received her B.E. from the P.B College of Engineering and completed her M.E. at the Anand Institute of Higher Technology. She obtained her Ph.D. from the Sri Sivasubramaniya Nadar College of Engineering (Anna University). Her research area includes image processing, signal processing, machine and deep learning.

Gajalakshmi PALANISAMY is an Assistant Professor in the Department of Computer Science and Engineering at the CK College of Engineering and Technology, Cuddalore. She completed her B.E. Degree at Arunai Engineering College and her M.E. Degree at Annamalai University. She has more than 10 years of active teaching experience in the field of computer science and engineering. Her areas of specialization include digital image processing and pattern recognition. She has been actively involved in academic research and has published several research papers in these domains. Her teaching and research interests focus on applying advanced computational techniques to solve real-world problems.