

Application of New Feature Techniques for Multimedia Analysis in Artificial Neural Networks by Using Image Processing

Lianqiu Liu^{1*}, Yongping Yang², Hong shun Chen²

¹School Of Artificial Intelligence and Big Data, Chongqing College of Finance and Economics, Chongqing, China, 402160

²Department of Quality Monitoring and Development Planning, Hunan University of Humanities, Science and Technology, Loudi, Hunan, 417000, China

E-mail: cqcf007@163.com

*Corresponding author

Keywords: kernel principal component analysis (KPCA), multimedia analysis, artificial neural networks (ANN), image processing

Received: March 6, 2024

To evaluate and extract information from multimedia material including photos, videos, and audio, a fusion of image processing and computer vision methods known as kernel principal component analysis (KPCA) is used. The objective is to create a system that can automatically identify relevant aspects of multimedia data and make them available for analysis and decision-making. By merging various processing methods, the fusion of image processing and multimedia analysis may improve analysis efficiency. The CT scans dataset from Kaggle is used to train artificial neural networks (ANN) for use in multimedia analysis. For multimedia analysis by combining image processing, we proposed kernel principal component analysis with artificial neural networks (KPCA-ANN) in this paper. The ability to process, analyze, and understand multimedia data by merging image processing into multimedia analysis has tremendous promise. It may improve decision-making, deepen our comprehension of complicated processes, and provide more fruitful means of information exchange. The experimental findings demonstrate that the proposed strategy has provided an absolute mean error of 7 and a structural similarity index of 86.

Povzetek: Predlagana metoda z uporabo KPCA-ANN izboljša analizo večpredstavnostnih podatkov z ekstrakcijo skritih značilnosti in natančno napovedjo, kar povečuje učinkovitost in kakovost odločanja.

1 Introduction

In recent years, multimodal image fusion has risen to prominence in the domain of medical technology. Single pieces of medical equipment can only produce single-modal images, which results in restricted information. To provide an accurate diagnosis, doctors frequently require a large number of multimodal pictures. Direct use of multimodal pictures increases every burden associated with illness diagnosis and increases the likelihood of interference and incorrect diagnoses. Medical professionals have found that fusion algorithms are a great way to combine the vast amounts of data included in multimodal pictures. Multimodal medical imaging is increasingly used, with MRIs and CT scans being the most common examples (CT) [1]. The evolution of culture and civilization the development of e martial arts, and illustrious history. As a

culture, they are handed down from generation to generation. In addition to being a symbol of athletic competition, it is also a symbol of traditional culture. While communicating with members of the international community, a nation needs to project a positive picture of itself. Every nation typically possesses some form of culture to serve as an image representative of national interactions [2]. According to statistics, the yearly death, incidence rate of brain cancer is a third higher than that of stomach cancer, lung cancer, and lung cancer combined. Each year, over 80,000 individuals lose their lives to brain cancer, which accounts for approximately 22.5% of the total number of cases of brain cancer worldwide. The number of persons being diagnosed with brain cancer has been steadily rising throughout the last few decades of the 21st century, particularly in Asian nations [3]. Because of limitations in

imaging technology, impossible to obtain a picture in which all subjects are sharply focused at the same time. To be more specific, in the case of an optical or convex lens with a fixed focal setting, objects be caught in focus if they fall within every low Depth-of-Field (DOF), though other things that are outside of DOF would be blurry [4]. As compared to English, written Chinese lacks both natural and explicit word borders and separators between words. The recognition of entities in Mandarin can be accomplished in one approach by first breaking down sentences into word sequences and then using those word sequences as the basis for entity recognition [5]. Another promising field for neural networks to play important roles in addressing issues and offering answers is medical imaging, and this field has a lot of potential for further development [6] [7]. With the ongoing advancement in both information processing technology and medical imaging technologies, a wide variety of medical pictures are now available for use in clinical diagnostics. Many imaging concepts and imaging technologies are applied in the creation of the various modalities that make up medical imaging. These medical images illustrate a variety of facts about the same human tissues and organs; nonetheless, they were created for medical purposes. Computed Tomography (CT) scans, which have a spatial resolution of less than one millimeter and are utilized extensively in a wide range of clinical illness assessments and aided diagnostics, can offer clear anatomical information about human bone structure [8]. The expression image enhancement refers to the process of boosting the visibility of specific information included within an image following a particular requirement, all while reducing any information that is regarded to be superfluous and enhancing the image's overall quality. The goal of image enhancement is to draw greater attention to particular aspects of an image so that the treated picture will be more suited for both human visual qualities and computer analysis, allowing for more sophisticated forms of image processing and analysis to be performed [9]. A strong Artificial Neural Networks (ANN) will need to be trained on a broad dataset to guarantee that it can properly analyze all different sorts of images. In general, although ANNs have the potential to significantly enhance brain CT image processing, it is essential to thoughtfully assess the possible drawbacks and restrictions of technology [10].

The paper's [11] goal of info concealing is to include a significant number of sensitive information stored in visuals, including text, audio, rigid and moving images, and motion. For penetration testing, image-based content concealing has been a crucial subject. Active picture deep

embedding techniques have been proposed in this case to conceal data. To incorporate a hidden signal into the initial image, Lowest Substantial Bits (LSB) stenography, a data encrypting technique, is recommended. The paper [12] separates images in the interactive media advanced embedded equipment into the benefits of the minimum and maximum components using the principal component analysis, and then develops a representation for the multimodal sequencing methods based on the Gabor filter. The integrand is best when four levels of spectral analysis are used, and when the cubic b-spline quantized is utilized as the quadratic target arrangement, the prediction accuracy is 89.08%. The paper [13] to better analyze data from multimedia, this work tries to enhance the pattern recognition version. Several methods undergoing validation set to identify the pattern may be used to put this into practice. One of these methods uses the Lexicon Index (LSI) and Euclidean distance (VSM) to use in textual data based on the type of query syntax. The alternative approach recovers similar qualities using wavelet segmentation and statistical metrics like mean, sample variance, and signal energy. The paper [14] provided a thorough grasp of the mass spectrometry used in the domain of visual recognition, to highlight the most significant developments, and to provide insight into potential future research. The connection between pattern recognition and image processing problems further, we may learn more about what makes deep learning effective as well as develop new deep models and training techniques. The paper [15] discipline of vision and image processing in computers has extensively used image individual works. Study [16] evaluates three major algorithms Single Shot Detection (SSD), You Only Look Once (YOLO), and Faster Region-based Convolutional Neural Networks (Faster R-CNN) for detecting images. As a finding, YOLO and SDD had better performance among the three algorithms based on parameters. The paper [17] examined important characteristics of sensory perceptions that are necessary for intelligent multimedia processing. Every goal is to demonstrate neural networks are an essential component of the mentioned multimedia functionalities: effective audio/visual intelligence generalizations, detection and classification methods, a fusion of multimodal signals, and multimodal regeneration and synchronization. Moreover, it shows how adaptive neural network technology offers a uniform response to a wide range of multimedia applications. Pertinent instances of neural networks being effectively applied to intelligent multimedia processing applications are noted as factual information. This research

[18] offers a neural network-based method for multimedia assurance. Artificial neural networks are most appropriate for this task because they have qualities like excellent security, little variation, and the capacity to operate for asymmetric input parameters. This is because encryption is a one-way functioning. The approach has applications in battlefield image database sharing, medical imaging techniques, and private streaming video, among other fields. Study [19] provided a novel data-enriched-transfer learning framework that consists of an support vector machine (SVM)-tuned trained VGG16 model with enhanced training data. The initial use cases for testing and fine-tuning are two

Twitter image collections. In the findings, SVM assures the highest levels of total accuracy and recall. In the paper [20] at the beginning, just one manually altered picture was used to identify any object fiddling, however in-universe, a real individual may be altered using a variety of image manipulation methods. It is now more difficult for a detector to identify tampering since numerous tampering actions are conducted on images and post-processing is used to remove the traces left behind. Newer approaches for detecting picture alteration rely on convolutional neural networks. Table 1 represents the summary table of related research papers.

Table 1: Summary table

Reference	Key metrics	Method	Findings
[11]	Accuracy	It was suggested to employ the least significant bit (LSB) steganography technique to encrypt an original image with a secret message.	The experiment’s outcome demonstrates the model’s resilience in terms of efficiency when compared to other cutting-edge systems with an accuracy rate of 95.1%.
[12]	Accuracy	Separating the image into the advantages of the high-energy and low-frequency components used the four-layer wavelet transformation	Applying four layers of wavelet decomposition yields an optimal categorization result with an accuracy rate of 89.08%.
[13]	training time and precision curve	ANN to enhance the multifaceted data retrieval search model	The neuronal network-based curvelet has a longer time elapsed than neural networks based on DWT or histogram, according to the training phase findings.

[16]	Accuracy, F1 score, and precision	Single Shot Detection (SSD), You Only Look Once (YOLO), and Faster Region-based Convolutional Neural Networks (Faster R-CNN).	YOLO and SSD are the best algorithms of these three algorithms.
[18]	Encryption and decryption	An approach based on neural networks for multimedia assurance and cryptosystem with neural network.	A multi-layered neural network's training approach is accessible for maintaining multimedia data securely.
[19]	Accuracy and recall	Data-augmented-transfer developed architecture, which is based on a VGG16 algorithm that has already been trained and improved used SVM with more training data.	According to the findings, VGG16-SVM generates the best overall accuracy (94%) and recall (96%).

2 Materials and methods

In this study, we suggest using kernel principal component analysis with artificial neural networks (KPCA-ANN) for multimedia analysis using image processing. The large dataset was collected for the TianChi contest to train ANNs for application in multimedia analysis. KPCA is a popular method for extracting features from images. Our planned research procedure is shown in Figure 1.

2.1 Dataset

We gather CT scan datasets from Kaggle (<https://www.kaggle.com/datasets/trainingdatapro/compute-d-tomography-ct-of-the-brain>). We utilize 80% data for training and 20 % data for testing. High accuracy rates were attained by the best models, which also showed promise for improving the early identification and diagnosis of brain disorders.

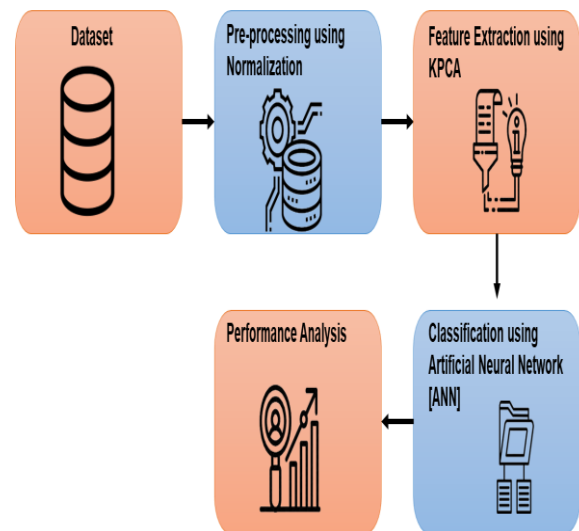


Figure 1: KPCA-ANN flow diagram

2.2 Pre-processing

In multimedia analysis, particularly in image processing, data normalization is a crucial pre-processing step. To make data easier to study and compare, normalization is the act of scaling or changing information into a common range or format. To enhance an image's visual quality and make it simpler to extract features, normalization in image processing may be used to alter an image's brightness, contrast, and color levels. In the field of image processing and multimedia analysis, min-max normalization is widely used. With this method, the pixel values of an image are transformed into a new range, often between 0 and 1. To do this, we first calculate the range of pixel values by subtracting the lowest pixel value from each pixel in the image and then dividing that number by the maximum and minimum pixel values.

Min-max normalization is a technique used to guarantee that all images in a dataset have the same range of pixel values, which is useful for some different kinds of multimedia research. When analyzing photos using machine learning algorithms, for instance, normalizing the input data is crucial for removing potential sources of error. In addition, min-max normalization may enhance an image's contrast and dynamic range, which makes it easier to select critical details.

The min-max normalization formula is as follows:

$$pixel_{new} = \frac{(pixel_{old} - min_{pixelvalue})}{(max_{pixelvalue} - min_{pixelvalue})} \quad (1)$$

Where: $pixel_old$ represents the image's original pixel value; min_pixel_value denotes the image's minimum pixel value; max_pixel_value denotes its maximum pixel value; and $pixel_new$ represents the image's new, normalized pixel value. To create the normalized image, this equation is applied to each pixel.

2.3 Feature extraction using KPCA

KPCA's fundamental notion is simple and familiar. In most cases, a linearly variable data collection is required for PCA to provide accurate results. The data may be transformed into a linear space of greater dimensions whenever the fluctuations are nonlinear. In other words, Cover's theorem states that high-dimensional nonlinear mapping often results in a linear data structure in input space creating a nonlinear relationship from the input area to the characteristic space using a simple kernel-based function, KPCA provides a computationally tractable solution. The input space

undergoes a nonlinear PCA operation, which KPCA conducts.

Each image must be depicted as a high-dimensional vector of features to use KPCA for image processing. If a PCA is used to diagonalizable the covariance matrix of a given collection of data, such as $x_k \in R^m, k=1, \dots, N$, to decouple nonlinear correlations between them, Instead of using an exponential input space, the coefficient of variation can be expressed in a linear characteristic space F., as shown in equation

$$C^F = \frac{1}{N} \sum_{j=1}^N \Phi(x_j) \Phi(x_j)^T \quad (2)$$

Where it is assumed that input vectors are projected from input space to F by the nonlinear mapping function $\Phi(\cdot)$ and that $\sum_{k=1}^N \Phi(x_k) = 0$. Every eigenvalue issue in feature space must be resolved to diagonalizable the covariance matrix.

$$\lambda v = C^F v \quad (3)$$

Where the eigenvalues are $v \in F \setminus \{0\}$ and $\lambda \geq 0$. The first PC in F is formed by the v and the greatest value given by equation (3), while the final PC is formed by the v and the lowest value. In this case, $C^F v$ may be written as (4)

$$C^F v = \left(\frac{1}{n} \sum_{j=1}^N \Phi(x_j) \Phi(x_j) \right) v = \frac{1}{N} \sum_{j=1}^N (\Phi(x_j), \langle v \rangle \Phi(x_j)) \quad (4)$$

Where the dot product of x and y is indicated by $\langle x, y \rangle$. This suggests that any solution with expression v with $\lambda \neq 0$ must fall within the range of $\Phi(x_1), \dots, \Phi(x_N)$. $\lambda v = C^F v$ is therefore identical to

$$\lambda \langle \Phi(x_k), v \rangle = \langle \Phi(x_k), C^F v \rangle, \quad k = 1, \dots, N \quad (5)$$

Furthermore, coefficients $x_i (i = 1, \dots, N)$ occur that

$$v = \sum_{i=1}^N x_i \Phi(x_i) \quad (6)$$

Equations (5) and (6) combined, we get

$$\lambda \sum_{i=1}^N x_i \langle \Phi(x_k), \Phi(x_i) \rangle = \frac{1}{N} \sum_{i=1}^N x_i \langle \Phi(x_k), \sum_{j=1}^N \Phi(x_j) \rangle \langle \Phi(x_j), \Phi(x_i) \rangle \quad (7)$$

The eigenvalue issue in equation (7) only concerns the dot products of mapped form vectors in feature space for all, $k = 1, \dots, N$. The mapping $\Phi(\cdot)$ exists but is often not always computationally tractable.

2.4 Classification using ANN

Biological neural networks in the brain serve as inspiration for ANNs, which are computer models. It's made up of stacked layers of interconnected processing nodes termed artificial neurons or just neurons. Neurons in a network operate by taking signals sent to them by other neurons, applying some kind of mathematical function to that data, and then sending that processed signal on to the next neuron in the network or outputting the data altogether. The image processing ANN is represented by Algorithm 1. Pseudocode representations about many neurons in input, hidden, and output layers are numInputNeurons, numHiddenNeurons, and numOutputNeurons, respectively. Training Data is a collection of annotated images that are used to teach the network. The training procedure is done numerous times. For each given input image, the desired output specifies what should be produced. By using CreateLayer (), a new layer of neurons with the given number of neurons is generated. With the help of the ConnectLayers () function, we can establish weighted connections between all of the neurons in two different layers of neurons. Input values for the forward layer may be changed using the setInputValues () method. The network's output values are determined by the propagate () function, the SetOutputError () function determines the error for each neuron in the output layer, and the Reverse () function reverses the forward error. The mistake is sent across the network through the Propagate () method, and the subsequent method Weights () modifies the neuronal connection weights. When the input image has been processed, the values for the output layer may be retrieved using the getOutputValues () method.

Algorithm 1: Artificial neural network

```

Start up the neural network

InputLayer = createLayer(numInputNeurons)
HiddenLayer = createLayer(numHiddenNeurons)
OutputLayer = createLayer(numOutputNeurons)
ConnectLayers (input layer, hidden layer)
ConnectLayers (hiddenLayer, outputLayer)

Train the neural network

For each epoch in numEpochs:

For each image in trainingData:
```

```
Forward propagation
```

```
SetInputValues (inputLayer, image)
```

```
ForwardPropagate (inputLayer)
```

```
Backpropagation
```

```
SetOutputError (outputLayer, desiredOutput)
```

```
backwardPropagate(outputLayer)
```

```
Update weights
```

```
updateWeights (hiddenLayer)
```

```
updateWeights (outputLayer)
```

```
Image processing using a neural network
```

```
SetInputValues (inputLayer, image)
```

```
ForwardPropagate (inputLayer)
```

```
Result = getOutputValues (outputLayer)
```

2.5 Hybrid of KPCA-ANN

The hybrid model that combines Artificial Neural Networks (ANN) with Kernel Principal Component Analysis (KPCA) offers a strong foundation for training and extraction of features. The feature extraction stage of this architecture uses KPCA to efficiently capture nonlinear relationships in the data. Simply the most informative main components are retained once the input data is transformed into a high-dimensional space using KPCA. The ANN phase, which normally consists of an input layer, many layers that are hidden, and an output layer, then receives these features that were extracted. The input layer serves as the neural network's first input by receiving the reduced-dimensional characteristics generated by KPCA. Convolutional layers of information and fully interconnected layers are two examples of hidden layer types that are used to learn intricate patterns and representations from data. This hybrid approach provides an adaptable structure for problems that include reducing dimensionality to regression and classification in a number of domains. It does that by utilizing the characteristics of both ANN for consistent training and KPCA to perform efficient feature extraction.

3 Result and discussion

The quality of the proposed KPCA-ANN strategy is thoroughly investigated through comparison and evaluation of the outcomes. A proposed method's precision and effectiveness are contrasted with those of contemporary methods like the Weighted Feature fusion of Convolutional neural network and Graph attention network (WFCG) [23], Convolutional Neural Network (CNN) [21], and Artificial Intelligence Fusion Model-Colorectal Cancer (AIFM-CRC) [22] to show that it is effective. The estimated Root Mean Square Error (RMSE), Peak Signal-to-noise Ratio (PSNR), Mean Absolute Error (MAE), Structural Similarity Index Measure (SSIM), and Entropy are shown in the results for the provided approach.

Peak Signal-to-noise Ratio (PSNR): calculates the similarity between reference picture and fused image and is defined as,

$$PSNR = 10 \times \log_{10} \left(\frac{255}{RMSE} \right) \tag{8}$$

High PSNR also denotes superior fused pictures, where RMSE is the root mean square error between reference and fused images, as shown in Figure 2.

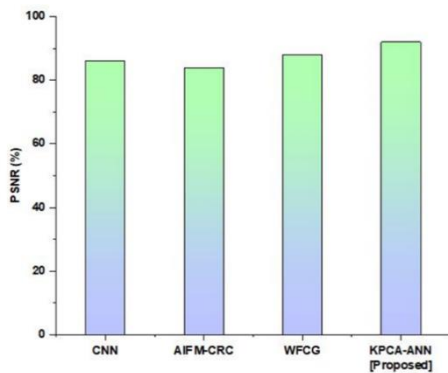


Figure 2: Peak Signal-to-noise ratio

The proposed method achieves the proposed 95% accuracy, CNN has obtained 86%, AIFM-CRC has gained 84%, and WFCG has scored 88%. Table 2 illustrates that the proposed strategy is much more effective than the existing one.

Table 2: Peak Signal-to-noise ratio

PSNR (%)	
CNN	86

AIFM-CRC	84
WFCG	88
KPCA-ANN [Proposed]	92

Root mean square error (RMSE) is calculating the square root of the average variance between the sort of disturbances that are expected and those that follow the equation:

$$RMSE = \sqrt{\frac{\sum_{u=1}^b (W_o - W_i)^2}{b}} \tag{9}$$

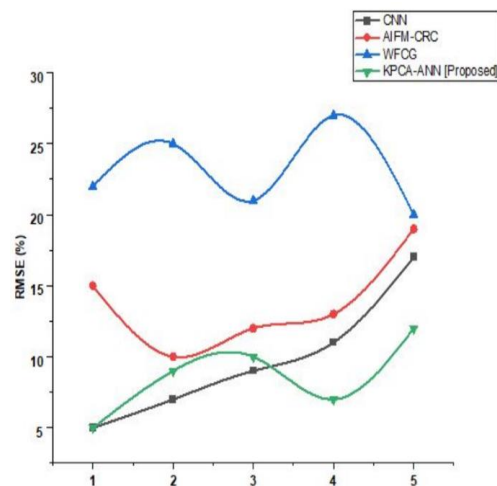


Figure 3: Root mean square error

Figure 3 shows the suggested system's Root Mean Square Error (RMSE). Every consumption prediction about Root Mean Square Error (RMSE) in existing systems and the proposed system is denoted. CNN has attained 17%, AIFM-CRC has acquired 19%, and FAWT has reached 19%, whereas the proposed system attains 12% of precision. It shows that the proposed approach is more effective than the existing one, as shown in Table 3,

Table 3: Root mean square error

RMSE (%)				
	CNN	AIFM-CRC	WFCG	KPCA-ANN [Proposed]
1	5	15	22	5
2	7	10	25	9

3	9	12	21	10
4	11	13	27	7
5	17	19	20	12

Entropy ϵ Calculates the amount of information in the image. It is defined by,

$$E = \sum_{j=0}^{1-1} p_j \log_2 P_j \tag{10}$$

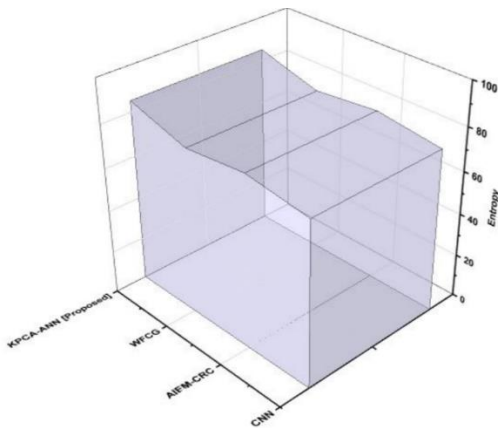


Figure 4: Entropy

Figure 4 shows the entropy, while the proposed method attains the proposed 85% accuracy, CNN has obtained 80%, AIFM-CRC has gained 80%, and WFCG has attained 77%. It illustrates that the proposed strategy is superior to the existing one, as shown in Table 4.

Table 4: Entropy

	Entropy
CNN	76
AIFM-CRC	80
WFCG	77
KPCA-ANN [Proposed]	85

Mean absolute error (MAE) is the difference between the classifiers anticipated and value also the actual value may be stated as in the equation that follows:

$$MAE = \frac{|\sum_u^b (w_o - w_i)|}{b} \tag{11}$$

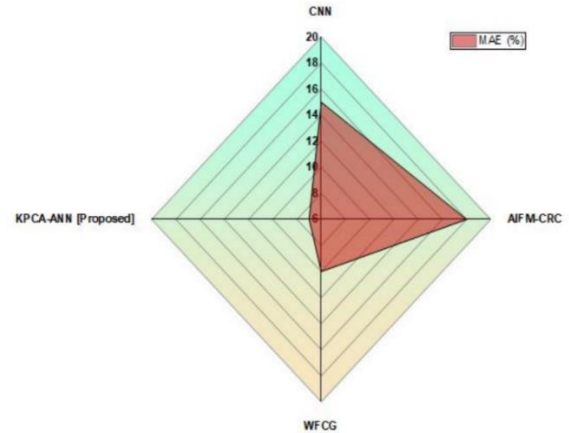


Figure 5: Mean absolute error

Figure 5 shows the Mean Absolute Error (MAE) of the proposed system. Where w_i represents the estimated value of the classifier in the supplied sample data and w_o is the observed value. The consumption prediction of Mean Absolute Error (MAE) in existing systems and the proposed system is denoted. CNN has attained 15 %, AIFM-CRC has acquired 18 %, and WFCG has reached 10 %, whereas the proposed system attains 7 % of precision. It shows that the proposed approach is more effective than the existing one, as shown in Table 5.

Table 5: Mean absolute error

	MAE (%)
CNN	15
AIFM-CRC	18
WFCG	10
KPCA-ANN [Proposed]	7

A popular metric for determining how similar two pictures are in image processing is the Structural Similarity Index Measure (SSIM).

$$SSIM(a, b) = [l(a, b) * c(a, b) * s(a, b)] \tag{12}$$

Where, $l(a, b)$ is the luminance factor, $c(a, b)$ is the contrast factor, $s(a, b)$ is the structural similarity factor, and x and y are the two pictures being compared.

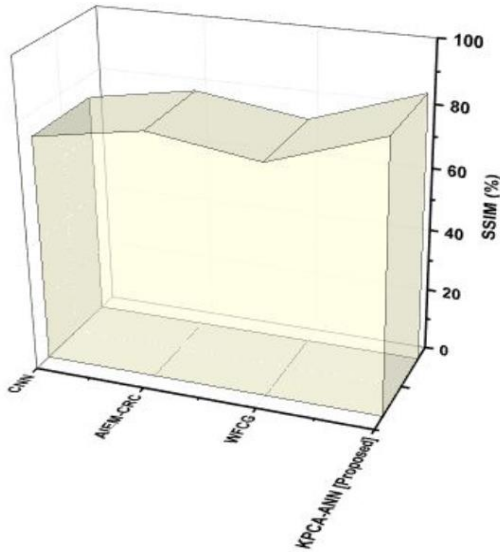


Figure 6: Structural similarity index measure

Figure 6 shows the structural similarity index measure, while the proposed method achieves the proposed 86% accuracy, CNN has attained 73%, AIFM-CRC has scored 79%, and WFCG has achieved 74%. Table 6 illustrates that the recommended approach of action is much more effective than the existing method.

Table 6: Structural similarity index measure

	SSIM (%)
CNN	73
AIFM-CRC	79
WFCG	74
KPCA-ANN [Proposed]	86

3.1 Discussion

Our proposed method was compared with existing methods including CNN, AIFM-CRC, and WFCG. CNN can be difficult to make an exact decision because it lacks clear comprehension. CNN attained PSNR (86%), RMSE (17%), Entropy (76%), MAE (15%) and SSIM (73%). The lack of

ability of AIFM-CRC to generalize to an extensive spectrum of CRC patients and its reliance on the diversity and quality of the input data, which could have an impact on how well the model performs in various patient populations or disease phases, are among its drawbacks. AIFM-CRC achieved PSNR (84%), RMSE (19%), Entropy (80%), MAE (18%) and SSIM (79%). The processing constraints of WFCG and its potential difficulty in handling complicated hyperspectral data pose limits that might hinder its scalability and real-time usage. WFCG achieved PSNR (88%), RMSE (20%), Entropy (77%), MAE (10%) and SSIM (74%). Our proposed method overcomes these limitations effectively and demonstrates superior performance such as PSNR (92%), RMSE (12%), Entropy (85%), MAE (7%), and SSIM (86%).

3.2 Practical implication

Improved diagnosis effectiveness and precision in hospitals are made possible by the results, which have important implications for medical imaging. Promising practical uses include modified treatment planning, disease tracking, and computer-aided diagnostics. All of these will eventually result in better patient outcomes and healthcare provision.

4 Conclusion

In this study, we propose using kernel principal component analysis with artificial neural networks (KPCA-ANN) to perform multimedia analysis using a combination of image processing. The aim of image analysis is to get useful data from images by breaking them down into their constituent parts. Images are a crucial information carrier because of their intuitive visuals and rich substance. They have a crucial role in conveying meaning. The ability to process images is now a staple of multimedia processing software. Multimedia analysis ANN trained on a big dataset collected from the TianChi contest. The experimental findings demonstrate that the proposed strategy outperforms the existing approaches. Scalability and appropriate representation of information may be limited by the computationally demanding nature of KPCA and ANN integration and the sensitivity to kernel function selection, which could restrict the generalization and interpretability of the suggested multimedia analysis system. In order to efficiently manage larger and more diversified multimedia datasets, future research might concentrate on strengthening the KPCA-ANN approach's adaptability and generalization, as well as its resilience to variability in data features.

References

- [1] Su, Y., Tian, J. and Zan, X., 2022. The research of Chinese martial arts cross-media communication system based on deep neural network. *Computational Intelligence and Neuroscience*, 2022.
- [2] Wan, Z., Dong, Y., Yu, Z., Lv, H. and Lv, Z., 2021. Semi-supervised support vector machine for digital twins-based brain image fusion. *Frontiers in Neuroscience*, 15, p.705323.
- [3] Li, J., Guo, X., Lu, G., Zhang, B., Xu, Y., Wu, F. and Zhang, D., 2020. DRPL: Deep regression pair learning for multi-focus image fusion. *IEEE Transactions on Image Processing*, 29, pp.4816-4831.
- [4] Gu, R., Wang, T., Deng, J. and Cheng, L., 2023. Improving Chinese Named Entity Recognition by Interactive Fusion of Contextual Representation and Glyph Representation. *Applied Sciences*, 13(7), p.4299.
- [5] Xu, Q., Zeng, Y., Tang, W., Peng, W., Xia, T., Li, Z., Teng, F., Li, W. and Guo, J., 2020. Multi-task joint learning model for segmenting and classifying tongue images using a deep neural network. *IEEE Journal of biomedical and health informatics*, 24(9), pp.2481-2489.
- [6] Hu, M., Zhong, Y., Xie, S., Lv, H. and Lv, Z., 2021. Fuzzy system based medical image processing for brain disease prediction. *Frontiers in Neuroscience*, 15, p.714318.
- [7] Gai, D., Shen, X., Chen, H., Xie, Z. and Su, P., 2020. Medical image fusion using the PCNN based on IQPSO in NSST domain. *IET image processing*, 14(9), pp.1870-1880.
- [8] Fu, J., Li, W., Du, J. and Xiao, B., 2020. Multimodal medical image fusion via laplacian pyramid and convolutional neural network reconstruction with local gradient energy strategy. *Computers in Biology and Medicine*, 126, p.104048.
- [9] Qiu, T., Wen, C., Xie, K., Wen, F.Q., Sheng, G.Q. and Tang, X.G., 2019. Efficient medical image enhancement based on CNN-FBB model. *IET Image Processing*, 13(10), pp.1736-1744.
- [10] Yu, Q., Shi, Y., Sun, J., Gao, Y., Zhu, J. and Dai, Y., 2019. Crossbar-net: A novel convolutional neural network for kidney tumor segmentation in ct images. *IEEE transactions on image processing*, 28(8), pp.4060-4074
- [11] Khan, A.A., Shaikh, A.A., Cheikhrouhou, O., Laghari, A.A., Rashid, M., Shafiq, M. and Hamam, H., 2022. IMG-forensics: Multimedia-enabled information hiding investigation using convolutional neural network. *IET Image Processing*, 16(11), pp.2854-2862.
- [12] Sui, K. and Kim, H.G., 2019. Research on application of multimedia image processing technology based on wavelet transform. *EURASIP Journal on Image and Video Processing*, 2019(1), pp.1-9.
- [13] Mahmood, M., Al-Kubaisy, W.J. and Al-Khateeb, B., 2023. Multimedia information retrieval using artificial neural network. *IAES International Journal of Artificial Intelligence*, 12(1), p.146.
- [14] Jiao, L. and Zhao, J., 2019. A survey on the new generation of deep learning in image processing. *IEEE Access*, 7, pp.172231-172263.
- [15] Xiang, H., Zou, Q., Nawaz, M.A., Huang, X., Zhang, F. and Yu, H., 2023. Deep learning for image inpainting: A survey. *Pattern Recognition*, 134, p.109046.
- [16] Srivastava, S., Divekar, A.V., Anilkumar, C., Naik, I., Kulkarni, V. and Pattabiraman, V., 2021. Comparative analysis of deep learning image detection algorithms. *Journal of Big data*, 8(1), p.66.
- [17] Kung, S.Y. and Hwang, J.N., 1998. Neural networks for intelligent multimedia processing. *Proceedings of the IEEE*, 86(6), pp.1244-1272.
- [18] Kumar, S.N., 2014. Technique for security of multimedia using neural network. Paper id-IJRETM-2014-02-05-020, *IJRETM*, 2(05), pp.1-7.
- [19] Jiang, Z., Zaheer, W., Wali, A. and Gilani, S.A.M., 2024. Visual sentiment analysis using data-augmented deep transfer learning techniques. *Multimedia Tools and Applications*, 83(6), pp.17233-17249.
- [20] Thakur, R. and Rohilla, R., 2020. Recent advances in digital image manipulation detection techniques: A brief review. *Forensic science international*, 312, p.110311
- [21] Bayar, B. and Stamm, M.C., 2016, June. A deep learning approach to universal image manipulation detection using a new convolutional layer. *In Proceedings of the 4th ACM workshop on*

information hiding and multimedia security (pp. 5-10).

- [22] Mansour, R.F., Alfar, N.M., Abdel-Khalek, S., Abdelhaq, M., Saeed, R.A. and Alsaqour, R., 2022. Optimal deep learning based fusion model for biomedical image classification. *Expert Systems*, 39(3), p.e12764.
- [23] Dong, Y., Liu, Q., Du, B. and Zhang, L., 2022. Weighted feature fusion of convolutional neural network and graph attention network for hyperspectral image classification. *IEEE Transactions on Image Processing*, 31, pp.1559-1572.

