

# Automatika

Journal for Control, Measurement, Electronics, Computing and Communications



ISSN: (Print) (Online) Journal homepage: [www.tandfonline.com/journals/taut20](http://www.tandfonline.com/journals/taut20)

## Multilayer vectorization to develop a deeper image feature learning model

D. Hemanand, N. P. G. Bhavani, Shahanaz Ayub, Mohd Wazih Ahmad, S. Narayanan & Anandakumar Haldorai

To cite this article: D. Hemanand, N. P. G. Bhavani, Shahanaz Ayub, Mohd Wazih Ahmad, S. Narayanan & Anandakumar Haldorai (2023) Multilayer vectorization to develop a deeper image feature learning model, *Automatika*, 64:2, 355-364, DOI: [10.1080/00051144.2022.2157946](https://doi.org/10.1080/00051144.2022.2157946)

To link to this article: <https://doi.org/10.1080/00051144.2022.2157946>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 31 Dec 2022.



Submit your article to this journal [↗](#)



Article views: 925



View related articles [↗](#)



View Crossmark data [↗](#)



## Multilayer vectorization to develop a deeper image feature learning model

D. Hemanand<sup>a</sup>, N. P. G. Bhavani<sup>b</sup>, Shahanaz Ayub<sup>c</sup>, Mohd Wazih Ahmad<sup>b,d</sup>, S. Narayanan<sup>e</sup> and Anandakumar Haldorai<sup>f</sup>

<sup>a</sup>Department of Computer Science and Engineering, S.A. Engineering College (Autonomous), Chennai, Tamil Nadu, India; <sup>b</sup>Department of Electronic Instrumentation Systems Institute of ECE Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, Tamilnadu, India; <sup>c</sup>Electronics and Communication Engineering Department, Bundelkhand Institute of Engineering and Technology, Jhansi, Uttar Pradesh, India; <sup>d</sup>Computer Science and Engineering, Adama Science and Technology University Adama, Adama, Ethiopia; <sup>e</sup>Department of Information Technology, SRM Valliammai Engineering College, Kattankulathur, Chengalpattu, India; <sup>f</sup>Department of Computer science and engineering, Sri Eshwar College of Engineering, Coimbatore, Tamil Nadu, India

### ABSTRACT

Computer-Aided Diagnosis (CAD) approaches categorise medical images substantially. Shape, colour, and texture can be problem-specific in medical imagery. Conventional approaches rely largely on them and their relationship, resulting in systems that can't illustrate high-issue domain ideas and have weak prototype generalization. Deep learning techniques deliver an end-to-end model that classifies medical photos thoroughly. Due to the improved medical picture quality and short dataset size, this approach may have high processing costs and model layer restrictions. Multilayer vectorization and the Coding Network-Multilayer Perceptron (CNMP) are merged with deep learning to handle these challenges. This study extracts a high-level characteristic using vectorization, CNN, and conventional characteristics. The model's steps are below. The input picture is vectorized into a few pixels during preprocessing. These pixel images are delivered to a coding network being trained to create high-level classification feature vectors. Medical imaging fundamentals determine picture properties. Finally, neural networks combine the collected features. The recommended technique is tested on ISIC2017 and HIS2828. The model's accuracy is 91% and 92%.

### ARTICLE HISTORY

Received 28 June 2022  
Accepted 8 December 2022

### KEYWORDS

Multi-layer vectorization; CNN; image feature extraction; Coding Network Multi-layer Perceptron (CNMP)

## 1. Introduction

Image identification by computer programmes has grown in popularity and activity in machine learning and application-specific research due to the tremendous progress in digital image recording and storage technologies [1]. Fast and accurate annotation of medical images is increasingly important in Computer-Aided Diagnosing (CAD) systems, paving the way for an intelligent CAD system across most medical fields. A large number of Americans get a skin cancer diagnosis each year [2]. If it had been discovered sooner, several lives might have been spared. Numerous research publications have been published in the area of medical image classification. The focusing area, contrast, and white balance of these photographs obtained from diverse sources could differ. Furthermore, inner structures with various textures and pixel densities are commonly seen in these photographs. It is difficult to accurately define certain groups if the traditional attributes are used to describe medical images [3].

The numerical attribute space in which machine learning algorithms operate is a two-dimensional array with input rows and columns that include featured

characteristics and actual examples. Consequently, the image pixel must first be translated into vector representations to execute numerical machine learning on photos. This method, the first step in examining natural language processing, is known as vectorization. The geometrical primitives use vector graphics to depict a raster image. Because of these primitives, it is very small, scalable, editable, independent of resolution, and even smaller in file size. Due to their qualities, they are also appropriate for mobile apps. The vectorization process makes new possibilities for using vector images and their inclusion in the creative arts, transforming raster images into vector representations [4].

Deep learning has been the most active area of study in computer-based applications in recent years. Due to recent advancements in several academic domains, deep learning has been attempted to handle non-medical images. The deep model's design was originally explained by Hinton et al. [5]. Various deep models are created to overcome image issues. A deep CNN trained to recognize images was used in the Large-Scale ImageNet Visual Recognizing Challenges 2010 [6]. Because of this fruitful study, much work has already been put

into using this unique concept to address the issues associated with categorizing medical images.

The classification of clinical photographs into different categories to assist doctors in diagnosing illnesses and research is the most urgent topic in image identification. Two categories make up the categorization of medical imagery. The first step is to extract the image's relevant components. The next step involves creating models that classify the image database using the attributes. It used to be a difficult, time-consuming procedure for doctors to obtain properties from medical images and then classify them using their expert information. This approach is probably going to provide inconsistent or unpredictable outcomes. Based on prior research, the application study of medical image classification is quite significant. Numerous published works on this topic result from the researchers' efforts. It is now unable to carry out this duty successfully, nevertheless. After completing the categorization procedure, the data will help physicians identify illnesses that need further research. Therefore, figuring out how to accomplish this task is crucial.

Before the development of deep structures, several prior studies [7–9] used shallow approaches for categorizing images that only relied on their integration of shape, colour, and texture data. Low-level features, which are underrepresented in high-level issue domain theories and have limited generalizability, are generally cited as the primary problem with these models' feature extraction techniques. In the non-medical imaging industry, deep architectures have a lot of expertise [10–12]. Deep learning techniques are the extremely exciting part of machine learning, which require a quick approach to create a complete model that computes final classification labels from input image pixels. Because large datasets are necessary for deep networks with vectorization to gain significant properties. Medical databases usually have a shortage of photos since getting them is notoriously difficult. To directly handle a small dataset, a deep model has been used. In addition to these problems, it is shown that the model's interpretability is weak, and training this prototype generally involves a significant quantity of processing. A new model combines convenient features with deep networks to instantly retrieve high-level features for medical image classification. This model does not fully benefit from conventional doctors' experiences, but it does address the issues with conventional methods compared to deep models.

## 2. Contribution

This paper's primary contributions are reviewed as follows:

- (i) High-level characteristics are used with common features to categorize medical photos. A multilayer

vectorization is used in the pre-processing step to several pixels being trained using the deep CNN referred to as the coding network to recover high-level detailed aspects of the image rather than the suggested domain-transferring convolutional neural networks (DT-CNNs). The suggested model's performance and interpretability might be improved by including common aspects of medical images.

- (ii) There are two alternative methods for combining high-level and standard features. The typical procedure is cumbersome, time-exhausting, and difficult to execute. Therefore, one result is to offer a fixed argument representation among high-level and standard features. Another approach is presented to deal with these issues: a revolutionary design that cannot incorporate features but may nevertheless change their dimensions on its own.

## 3. State-of-the-Art

Two sorts of techniques – standard systems and deep model systems – are offered to handle these challenging image classification problems. Support vector machines and random forests are a few examples of traditional methods, as are colour and texture [13–15]. We'll begin by providing a thorough summary of earlier studies on image classification. Feature integration for image categorization issues will be assessed in the literature [16–18].

In the suggested random forests utilizing single-photon emissions computerized tomography (SPECT), the image categorization is carried out [19] to assist in the diagnosis of Alzheimer's disease (A.D.). They started by using partial least squares to obtain score features from datasets. The method regressively classifies the image to the closest centroid until the classified image is obtained. Incremental learning is the key component of this approach that can be understood from earlier models without needing to collect images from nothing.

A tailored CNN is created to classify images of lung patches. Compared to SIFT includes, rotating-invariant localized binary pattern features and unsupervised feature extraction using the restrained Boltzmann system, this model employed a single convolutional layer to obtain detailed features and had the best classification result (RBM). Principle component analyzing network (PCANet), a fundamental deep learning technique created by the researchers, has been used by [20] in combination with colour image spatial dispersion data to achieve cutting-edge classification accuracy across several databases [21]. To identify different types of illnesses in chest X-ray images, the researchers employed a CNN trained using ImageNet. They were able to get the greatest accuracy results by combining information pulled from CNN with tailored features. It is

detailed how medical imaging is used to convey learning. They also showed how their discoveries might be used to diagnose interstitial lung disease and identify thoracoabdominal lymph nodes (L.N.s) (ILD). Initially, localized binary patterns (LBP) and localized quinary patterns (LQP), which are robust to minute deflections during lung cancer diagnosis, were combined with the scattering transformation to extract features [22, 23]. They also evaluated the 2D-Hela and Pap smear performance effectiveness of the useful database. The whole sliding images (WSI) of breast cancer were validated using a dataset of 160 patients with infective ductal carcinoma (IDC) and reached balanced precision of 84.23 percent [24]. A deep learning approach is proposed for automatically recognizing IDC regional tissue.

A broader road vectorization approach is created by fusing and improving the methods described in their earlier work [25]. Utilizing interactive CIS, road pixels are being carefully extracted. The researchers developed a single-pass parallel sequence tracing algorithmic strategy for identifying road width and format, increasing the temporal complexity of the parallel pattern tracing approach. Road centerlines were then produced using morphological procedures based on the chosen format and width. Since the operators of the thinning bend the routes at crossings, the method examines the road routes inside the impacted regions. At road junctions, location upgrades have been made using the traced paths. To retrieve the vector information, the road centerlines are then traced using the exact coordinates of the road intersections. This approach depends on the diversity of road features, for vectorization is a disadvantage. Because their breadth is smaller than the width of the bigger road, some small road connections are accidentally demolished.

Line drawings are vectorized. Initial vectorization techniques identified geometric primitives, such as ellipses, straight lines, Bézier curves, and B-splines [26–28]. Line illustration vectorization is created by converting scanned technical drawings into electric schematics. Recognition-based algorithms can generate compact parametric curves using fitting techniques, but the vectorization's robustness seems dubious. For instance, extra treatments are necessary to avoid overfitting errors such as layer segregation when the lines in the input image have varied thicknesses.

#### 4. System model

Since the complete presentation of LeNet-5 has high efficiency, CNN is commonly applied in image categorization, object recognition and video surveillance. CNN's often have convolutional, pooling, fully connected, and softmax layers. Convolutional and pooling layers are referred to as the softmax layer for feature extraction. The Convolution Layers are the first layers

used to extract characteristics from an image. It maintains the link between pixels by learning characteristics from a small set of input data. It is a mathematical concept with two inputs: an image matrix and a kernel or filter. Using receptive fields, the convolutional layer detects (many) patterns in different sub-regions of the input field. Padding is essential in the development of CNN. The image's initial size is reduced after the convolution technique. In addition, there are numerous convolution layers in the image classification task, and our original image is shrunk after each step, which we do not desire. Second, as the kernel goes over the original image, it passes through the central layer more than the edge layers, causing an overlap. To address this issue, a new idea known as padding was established. It is an additional layer that can be added to the boundaries of an image while maintaining the original image's size. The pooling layer is another fundamental part of a CNN that is essential in image pre-processing. If the image is too large, the pre-process compresses it by reducing the number of parameters. The pixel density is reduced when the image is shrunk, and the downscaled image is produced from the previous layers. The pooling layer reduces the spatial size of the representation, the number of parameters and the amount of processing in the network, and hence controls overfitting. The pooling layer summarises the features in a region of the feature map produced by a convolution layer. Max pooling is a rule that takes the maximum of a region and helps to proceed with the image's most important elements. It is a sample-based method for converting continuous functions to discrete equivalents. Its major goal is to downscale an input by lowering its dimensionality and rejecting assumptions about characteristics included in the sub-region. Soft-max function to classify an object with probabilistic values 0 and 1. The softmax function is chosen as the activation function in the output layer of neural network models that predict a multinomial probability distribution. The proposed primary design principles include: first, image processing is carried out, in which the image is vectorized into several 2D pixels by removing the mean ZCA and RGB blanching; next, the appropriate stimulation function is chosen; and finally, initial weights are determined. When the initial weights appear too small or too large, the proposed deep network cannot learn, and the initial weights will diverge. Augmenting data also involves collecting random patches after the input image and flipping it horizontally, which is significant in medical image processing. Dropout is also used to prevent overfitting, and local normalized response reduces error rates. Decide on the best information gain. The suggested multilayer vectorization and classification architecture have been used to accomplish the most common practice, which is for the information gain to decline with each epoch. The following

method is used to extract the detailed image feature information.

#### 4.1. Image processing and vectorization

The pre-processing step comprises elimination and binarization based upon the organization of the input image. The source input image's red, green, and blue channels are typically combined into one channel and normalized. The vectorization of the image is then carried out. The image is separated into several pixels, and each pixel is independently examined using CNN networking to more precisely obtain the information.

For the multilayer vectorial portrayal of the source input image, I include the following layer during the pre-processing step.

- A supportive domain  $D$  covers an input image's pixels subset.
- $C(p)$  is a colour gradient function that assigns an RGB colour to each pixel in the image;  $p \in D$ .
- $A(p)$  defines the functional opacity gradient that assigns a value of opacity per pixel  $p \in D$ . If  $p \notin D_n$ , then set  $A(p) = 0$ .

Via recursive  $\alpha$ -blending of the ordered multilayers, the output image  $I_n$  is provided from an  $n$ -layer depiction.

$$I_n(p) = A_n(p)C_n(p) + (1 - A_n(p))I_{n-1}(p) \quad (1)$$

Whereas the formulation above is generic, the colour gradient and opacity gradient are limited to linear patterns with an identical orientation that conforms to the SVG format's 2 linear gradients. The colour vectorization includes  $C(p) = c_0 + c_1 O^t p$  as well as  $A(p) = a_0 + a_1 O^t p$  wherein  $c_0$  as well as  $c_1$  are colour vectors concerning opacity scalars ( $a_0, a_1$ );  $O$  represents the vector orientation. Though linear gradients became less descriptive rather than more sophisticated primitives as gradient combined and diffused curves, it is quite easier to manage due to their modest parameters. It is handled by nearly every vector graphics application.

#### 4.2. Coding network

The input medical image possesses a  $140 \times 140$  fixed-size RGB image, which initially undergoes the removal of the median RGB value per pixel by vectorization before transmitting the medical image to the coding network. A succession of pooling layers and convolutional layers make up the coding networking. A pixel of  $7 \times 7, 8 \times 8, 10 \times 10$ ; stride, and padding of 1 and 0 pixels, are being utilized in this convolutional layer. The following is the definition of the convolution operation:

$$y_j^r = f \left( b_j^r + \sum w_{i,j}^{r-1} * x_j^r \right) \quad (2)$$

Where the coding network with the  $r$ -th layer is represented by  $r$ , as well as the activation function has

**Table 1.** The coding network's comprehensive structure.

Type of Layer	Size of Patch/Stride	Resultant Size
Convolution Layer	$10 \times 10/1$	$132 \times 132 \times 30$
Convolution Layer	$10 \times 10/1$	$122 \times 122 \times 30$
Max pooling Layer	$4 \times 4/2$	$60 \times 60 \times 30$
Convolution Layer	$8 \times 8/1$	$52 \times 52 \times 62$
Max pooling Layer	$4 \times 4/2$	$25 \times 25 \times 62$
Convolution Layer	$7 \times 7/1$	$18 \times 18 \times 126$
Convolution Layer	$8 \times 8/1$	$10 \times 10 \times 254$
Convolution Layer	$7 \times 7/1$	$2 \times 2 \times 254$
Rasterize	-	$2 \times 2 \times 4$
Final Softmax Layer	-	$2 \times 2 \times 4$

been represented by  $f^r$ ; The  $i$ -th input, as well as  $j$ -th output feature maps, are represented as  $x_i$  and  $y_j$ ; the convolution kernel connecting  $x_i$  as well as  $y_j$  equals to  $w_{i,j}$ ; the bias has been represented by  $b_j$ ; \* represents convolution operational symbol

The pooling layer was created via max-pooling using  $4 \times 4$  windows and 2 strides, indicating an overlapping. The operational pooling is described as

$$y_{j,k}^i = \max_{0 \leq m, n < 4} (x_{j \cdot 4 + m, k \cdot 4 + n}^i) \quad (3)$$

Here, the  $4 \times 4$  input feature map  $x^i$  is locally overlapping to produce every component in the  $y^i$  output quality map. Softmax classifier is present at the final layer of the coding system to categorize the medical image based on the detailed vectorization to obtain the features. The coding network's comprehensive structure is shown in Table 1.

**Activation Function:** The functional sigmoid  $f(x) = (1 + e^{-x})^{-1}$  as well as the functional  $f(x) = \tanh(x)$  are the fundamental activation functions; its derivatives are expressed by themselves, and they may transfer the greater variation outcomes into a small distance. The two functions are similar issues of slower convergence rate; hence, there is a diffusion gradient issue.  $f(x) = \max(0, x)$ , Rectified Linear Units (ReLUs) functional activation of the coding network is computationally efficient and limit the diffusion impact of gradient. Furthermore, ReLUs are used to converge quicker in comparison with sigmoid or tanh.

**Softmax Layer:** In the proposed network, the final layer is linked with a softmax layer via which "n" various classes are predicted by estimating the probability of each classification. The feature is rasterized into  $x$ , where "x" denotes the column feature vector:

$$p(y = j|x, \theta) = \frac{e^{\theta_j^T x}}{\sum_{j=1}^k e^{\theta_j^T x}} \quad (4)$$

Here,  $\theta_j^T$  denotes weight vector; the objective goal comprises  $k$  classes.

Thus, the colour instant, as well as texture features, are employed as traditional features. Statistically, a texture distribution feature can be used to explain an



image's inherent features. Rather than single pixels, it uses many pixels area processing. The colour instant depends on a specific pixel and isn't highly reactive to the image's size or angle.

To estimate the texture features, the gray-level "G" co-occurrence matrix is obtained from the image to estimate the statistical pairs of neighbouring pixels. The angular secondary moment (ASM), entropy (ENT), contrast (CON), and correlation (COR) are used to represent the texture properties that may be extracted from the matrix G fixing the distance between one to two pixels with an angle of 0–135 degrees. The following are the descriptions:

$$ASM = \sum_i^s \sum_j^s G(i, j)^2 \quad (5)$$

Here, G represents the matrix of gray-level co-occurrence; s represents G's size; and  $G(i, j)$  denotes the row (i) and the column (j) component of the G matrix.

In matrix G, the angular secondary moment is defined as the addition of each square element. It represents the image's uniformity as well as the texture's hardness. The ASM value is minimal when there is an identity element in matrix G, which is huge if this does not happen.

$$ENT = - \sum_i \sum_j G(i, j) \log(G(i, j)) \quad (6)$$

Entropy is defined as an uncertainty measurement that can be utilized to represent the image's ambiguous data. The image includes the most quantity of unclear data depending on the maximum ENT value if all the components in matrix G remain the same. The distribution of grey-value in the image becomes extremely convoluted at this point.

$$CON = \sum_i \sum_j (i - j)^2 G(i, j) \quad (7)$$

Contrast measurement is involved in eventually spreading the data in an image and how clear the image appears. The image is viewed when the CON value is higher.

$$COR = \sum_i^s \frac{\sum_j^s ijG(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y} \quad (8)$$

Here,  $\mu_x$  and  $\mu_y$  represent the mean value of G's distribution whereas  $\sigma_x$  and  $\sigma_y$  represent the standard deviation of G's distribution.

The mean and standard deviation are estimated with each other after computing ASM, CON, ENT, and COR, resulting in a textural feature vector. With these traditional features, it will be combined with the colour instant.

The basic deviation, mean, and third-order colour instant must be used to illustrate colour features. The darkness or lightness of the image is obtained via mean; the image's colour dispersal range is obtained via standard deviation; the third-order colour moment will reveal the image's colour distribution symmetry. As a result, this method yields a colour momentary feature vector. The following is the description of a colour moment:

$$A_i = \frac{1}{N} \sum_j^N P(i, j) \quad (10)$$

$$V_i = \sqrt{\frac{1}{N} \sum_j^N (P(i, j) - A_i)^2} \quad (11)$$

$$S_i = \sqrt[3]{\frac{1}{N} \sum_j^N (P(i, j) - A_i)^3} \quad (12)$$

Here, P denotes the image's matrix representation; N denotes the number of pixels;  $A_i$ ,  $V_i$ ,  $S_i$  represent the mean, variance, and skewness of the input image's  $i^{th}$  channel and  $P(i, j)$  represents the pixel j in  $i^{th}$  channel.

**Feature Fusion:** Two alternative fusion methods are designed to integrate the features after isolating the high-level and traditional features. The first strategy, R feature fusion, is to specify fixed proportions. The following is how the categorization of integrated features is calculated:

$$NF = \lambda \cdot LF + (1 - \lambda) \cdot HF \quad (13)$$

Here, NF denotes the fusion feature; L.F. indicates the traditional features, whereas H.F. indicates the high-level features. The weighted parameter  $\lambda$  indicates the relative relevance of two separate qualities. Since it is globally weighted, this approach is quite simple to execute. There are zero requirements to recalibrate as the parameter  $\lambda$  is being obtained. Softmax used the integrated feature to complete the final classification process.

Furthermore, this method applies to linear feature integration and obtaining the parameter  $\lambda$  needed via a considerable series of experiments. Most importantly, fusing the features to portray the images adequately is tough. Furthermore, the identical experiment must be repeated when the dataset is changed to retrieve the parameter  $\lambda$ .

To address these issues, a new approach can autonomously alter the percentage of high-level characteristics to conventional features, avoiding the tedious and time-consuming procedure of parameter calculation. The method entails training a perceptron neural network that possesses multiple numbers of layers and is capable of fusing nonlinear space features. This

feature integration operation is as follows:

$$RF = \max \left( 0, \sum_i^n w_i l_i + \sum_j^m w_j h_j + b \right) \quad (14)$$

Here,  $LF = \{l_1, l_2, \dots, l_i, \dots, l_n\}$  representing the traditional features as well as  $HF = \{h_1, h_2, \dots, h_i, \dots, h_n\}$  representing the high-level features, respectively, and  $b$  denotes bias. In the case of feature classification, the multilayer perceptron has a fully connected layer followed by a softmax layer is utilized. The kernel functional concept aims to convert low-dimensional data into high-dimensional data. As a result, it can obtain more discriminative detailed features rather than linear space features. In the tests, this technique will be clearly illustrated. Furthermore, because it will not try to estimate the identical parameter many times, it might significantly minimize the computation amount.

## 5. Result and discussion

A Matlab toolkit, namely MatConvnet, creates multilayer vectorization and convolutional neural networks for constructing the coding network model to retrieve the high-level features and traditional features depending on texture and colour moment. Leveraging two standard medical imaging datasets, the ISIC2017 dataset and HIS2828 dataset, a series of tests are determined to evaluate the usefulness of our technique. All tests were performed on a machine using a 3.2 GHz CPU i5-6500, 32 G.B. of main RAM, and a GTX1060 GPU.

The HIS2828 dataset contains four different image types representing various tissue types, wherein each image possesses  $720 * 480$  RGB images. This dataset comprises 2828 images, 1026 nerve tissue images, 484 images of connective tissue, 804 images of epithelial tissue, and 514 images of muscle tissue. The International Skin Image Collaboration (ISIC2017) has produced a skin lesions dataset. It contains 2000 images, wherein 374 images of malignant skin cancers are termed “melanoma” while 1626 images of benign skin tumors are termed “nevus and seborrheic keratosis.” The binary input image vectorization and classification are quite challenging to differentiate between Melanoma and Nevus of Seborrheic Keratosis. This dataset is handled because each image has a different resolution.

The output quality of the multilayer vectorization is assessed in two ways. Because of primary concern with the visual component of vectorization of the input images from these two datasets, the output vectorized images are evaluated in terms of features is a fundamental concern of this research. The degree of detailed focus is determined by pixel vectorization. The categorization parameters are determined; however, the control point choice precise parameter of vectorization processes is

fixed to be (1/100). Using any classification approach with this vectorization allows for creating style ranges. (Figure 1)

Additional evaluation component, the image’s memory efficiency of vectorization methods is considered. Some regular sample medical images are vectorized using the appropriate parameters to provide visual outputs. Figure 2 illustrates that each layer’s vectorized medical tissue images have identical Peak Signal-to-Noise Ratio (PSNR) values. However, at this stage, the algorithm’s efficiency is determined by the bits-per-pixel (bpp) rate. Figure 3 shows the bpp ratings of various vectorized pixel tissue image datasets. The multilayer vectorization method ranks significantly higher in efficiency. This is owing to the established operational requirements of detailed medical image features in hospitalization applications.

Every dataset is classified under training, validation, and testing processes in 7:1:2 ratios. Then, all the approaches are tested utilizing 10-fold cross verification. The images are vectorized into many pixels from the actual dataset to create fixed size  $140 \times 140$  images fed as an input into the coding network. Every image of the HIS2828 dataset is randomized to  $420 \times 420$  pixels and then restored to  $140 \times 140$  image sizes. Before downsizing to  $140 \times 140$  for the ISIC2017 dataset, randomized patches are extracted with two-thirds of the actual image’s height and width of various resolutions. This saves a significant amount of image data while reducing processing difficulty. These works retrieve the fixed-size input images and supplement them. The image is flipped horizontally or vertically to enhance the image datasets even more. Simultaneously, the network provides an estimate per patch and averages the softmax layer’s estimation when the patches are from the identical image. Further, the augmented image impact on accuracy and its running time is addressed.

Table 1 shows the coding network topology in context. It is possible to converge following 45 epochs. Eventually, the ReLUs activation function is employed for each convolutional layer. Aside from that, batch normalization is used to speed up the deep network training. The accuracy rate and algorithmic running time on two medical image datasets are examined. The percentage of properly categorized medical images has been used to measure accuracy. The receiver operational characteristic (ROC) curve is used to assess the model and properly contrast the algorithms. The ROC curve represents a graphical representation that is created by comparing the true-positive rates (TPR) and the false-positive rates (FPR) using various thresholds, with TPR as well as FPR defined as follows:

$$TPR = \frac{TP}{TP + FN} \quad (15)$$

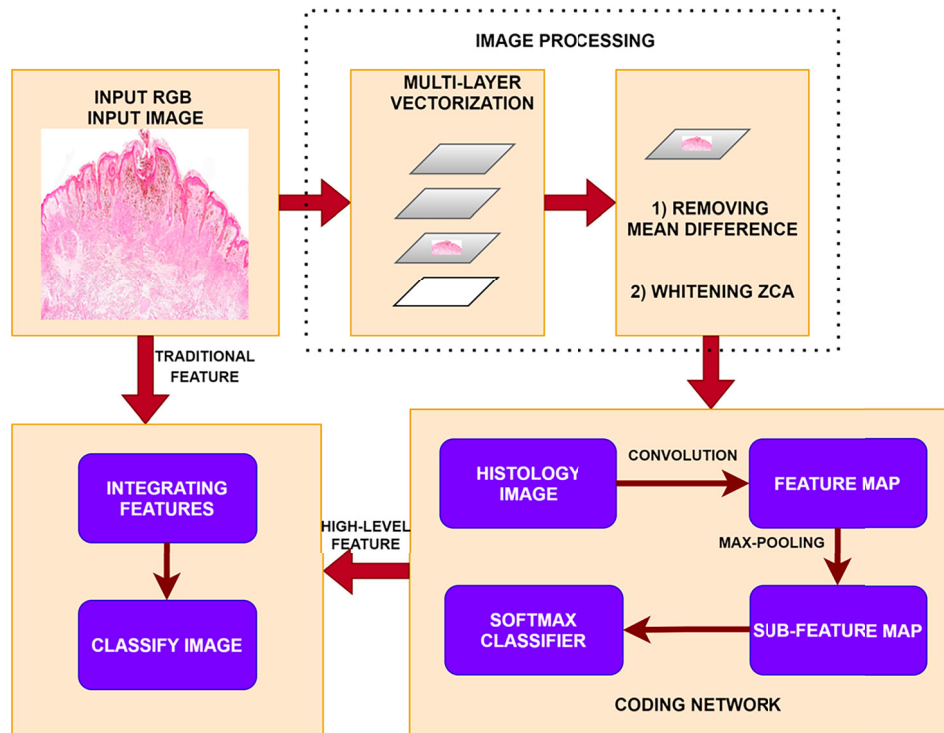


Figure 1. The proposed architecture model.

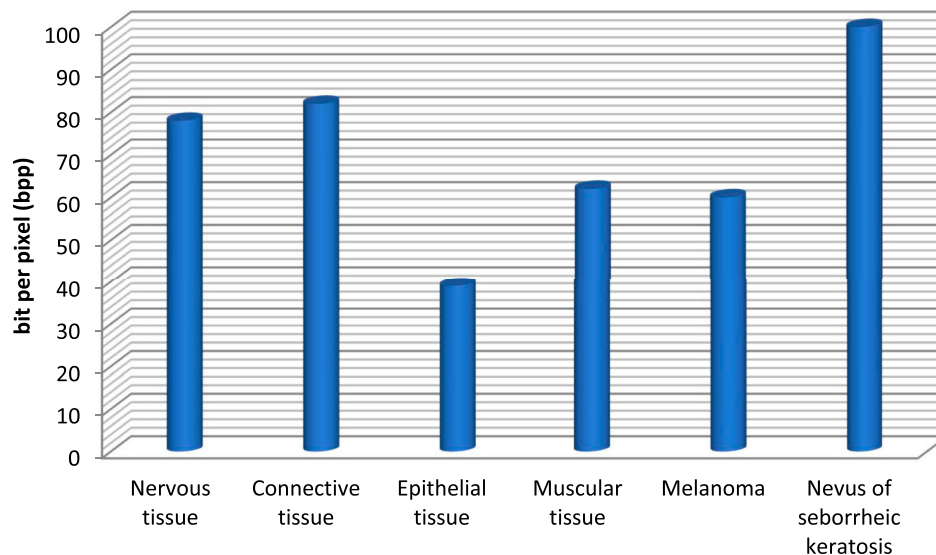


Figure 2. Bit per pixel (bpp) of various vectorized images.

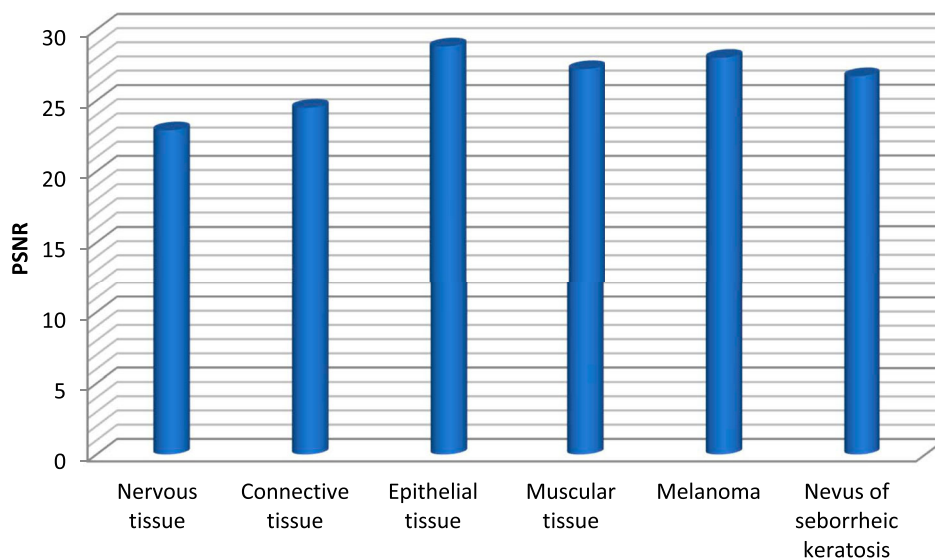
$$FPR = \frac{FP}{FP + TN} \quad (16)$$

Here, T.P. denotes true positive; F.P. denotes false positive value; F.N. denotes false negative, and T.N. denotes true negative. The image classification technique's performance is evaluated using it. Before the enhancement of the deep learning system, the SVM method is evaluated to be the preferable machine learning classifier; thus, the SVM concatenation of traditional and deep features is compared with the proposed CNMP model.

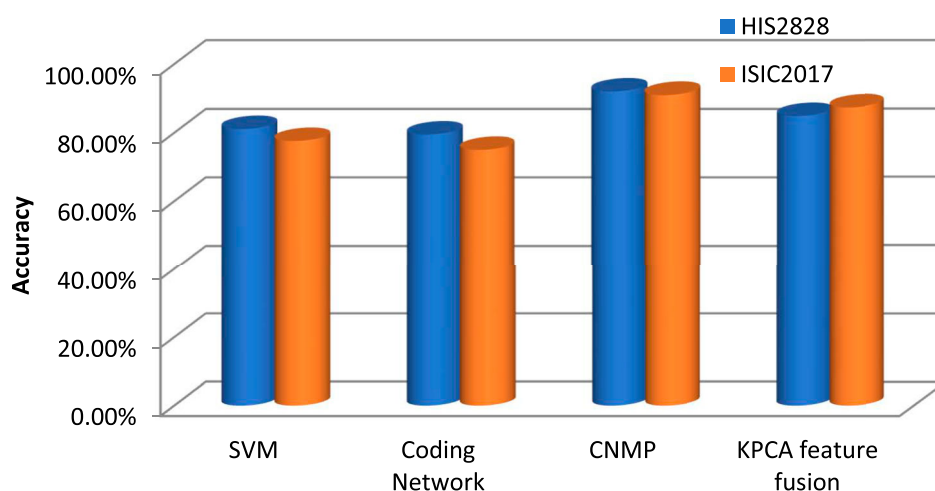
To train a multiclass classifier, radial basis functioning (RBF) kernel and LibSVM-3.17 library are utilized. The coding network requires illustrating the usefulness of integrating features. In addition, when contrasted to

KPCA feature integration, the CNMP has a superior feature integration strategy. KPCA includes the RBF kernel to combine features since it can trace the nonlinear feature space. The integrated feature vector will be sent into softmax to complete the categorization process. Figure 4 shows the test accuracy result on the HIS2828 dataset and the ISIC2017 dataset. The proposed technique gets 91% and 92% accuracy rates, correspondingly. The coding network is also used to categorize the medical image leads to better results. Thus, the medical images are better represented using high-level features than traditional features. SVM is superior to coding networks, and SVM contains traditional features. Furthermore, R feature integration and KPCA





**Figure 3.** PSNR of various vectorized images.



**Figure 4.** Classification accuracy of the various techniques.

feature integration are compared with the proposed model; the autonomous feature integration achieves superior results and eliminates the time-consuming procedure of manually modifying the parameters.

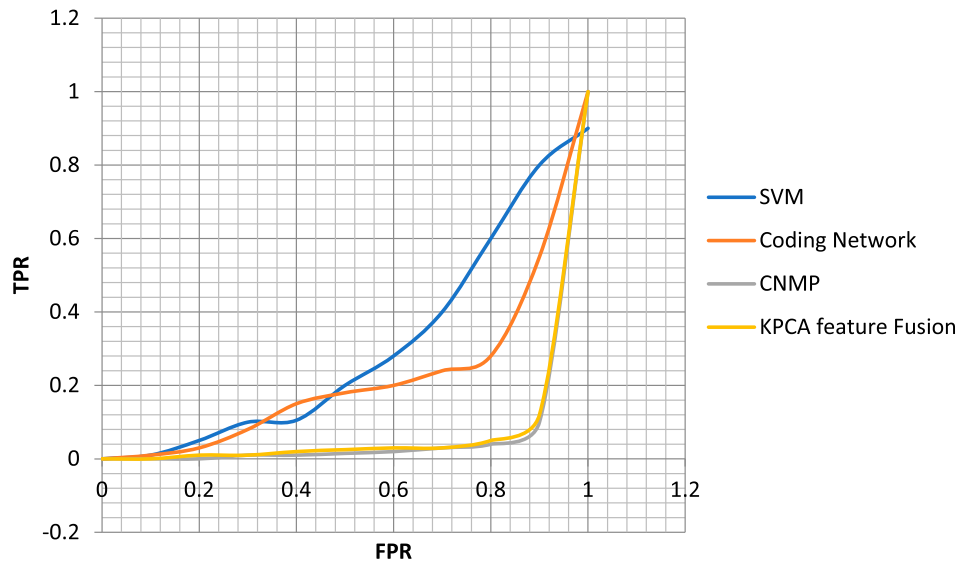
The ROC curve is represented for TPR and FPR derivation by varying the thresholds. Because the binary dataset includes an imbalanced sample problem, its classification technique's performance is evaluated. With this ISIC2017 dataset, the comparison of various techniques' ROC curves is illustrated in Figure 5, wherein the performance raises with curve closeness.

Figure 6 illustrates the comparison of various running times. SVM will run faster as the result of several causes. (1) It utilizes only a little quantity of image data while discarding an image's large portion of the spatial information. (2) Compared to the deep learning model, this method needs to train fewer parameters. The next quickest method is coding networks. However, it takes much longer to execute than SVM since the deep model needs to train many parameters to increase its generalization capabilities. Furthermore, combining

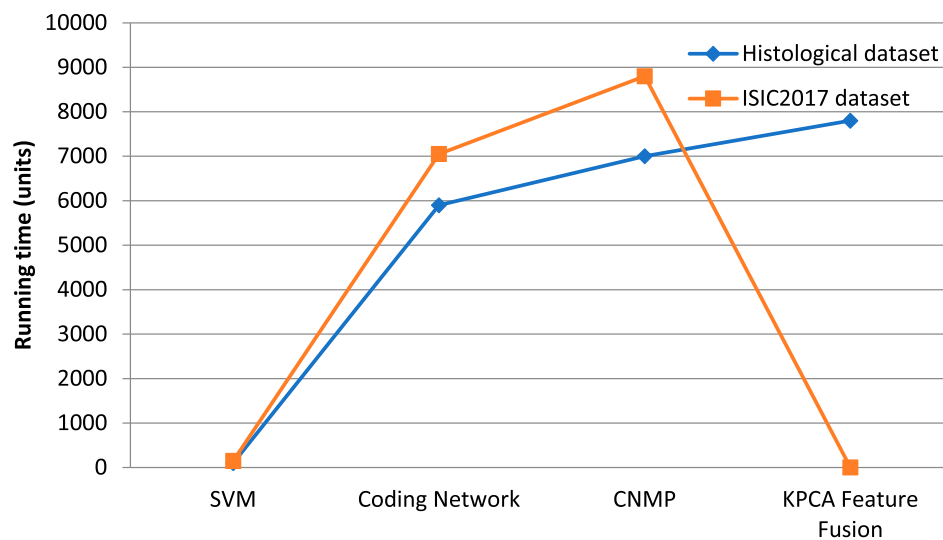
the multilayer features uses practically all informational features of an image. Moreover, CNMP has the highest accuracy; its running times are quite long. In addition, it must surmount excessive dimensionality to acquire the categorization model. The integration of the various aspects in the KPCA requires a long period.

## 6. Conclusion

This study uses a unique architectural model to extract the fine details from the input medical photos for accurate diagnosis and patient treatment. This model comprises the multilayer vectorization and classification of a medical image to get high-level features. The CNMP classification approach combines conventional properties of an input image with high-level extracted features from a coding network. The medical photos are categorized based on their specific visual attributes. The experiment results demonstrate how multilayer vectorization of tissues performs in terms of PSNR and bits-per-pixel on diverse datasets. The proposed CNMP



**Figure 5.** ROC curve obtained for various techniques.



**Figure 6.** Running time of various classification techniques.

outperforms coding network, SVM, and KPCA feature fusion by a wide margin in terms of classification accuracy, achieving a rate of 92 percent on the HIS2828 dataset and 91 percent on ISIC2017 image datasets. The approaches mentioned above' running times are also examined. This technique will eventually be used as a productive pruning strategy to reduce the parameters significantly. The multilayer vectorization makes the ability to extract images with greater accuracy.

### Disclosure statement

No potential conflict of interest was reported by the author(s).

### ORCID

Mohd Wazih Ahmad  <http://orcid.org/0000-0001-5614-2591>

### References

- [1] Cruzroa A, Caicedo JC, Gonzalez FA. Visual pattern mining in histology image collections using bag of features. *Artif Intell Med.* 2011;52(2):91–106.
- [2] Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature.* 2017;542(7639):115–118.
- [3] Zare MR, Mueen A, Awedh M, et al. Automatic classification of medical X-ray images: hybrid generative-discriminative approach. *IET Image Process.* 2013; 7(5):523–532.
- [4] Kozhevnikov VA, Pankratova ES. Research of the text data vectorization and classification algorithms of machine learning. *Theoretical Appl Sci.* 2020;85(85): 574–585.
- [5] Hinton GE, Osindero S, Teh Y-W. A fast learning algorithm for deep belief nets. *Neural Comput.* 2006;18(7):1527–1554.
- [6] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Commun ACM.* 2012;60(6):1097–1105.
- [7] Riaz F, Hassan A, Nisar R, et al. Content-adaptive region-based color texture descriptors for medical images. *IEEE J Biomed Health Inform.* 2017;21(1): 162–171.
- [8] Zhang Y, Chen S, Wang S, et al. Magnetic resonance brain image classification based on weighted-type fractional fourier transform and nonparallel

- support vector machine. *Int J Imaging Syst Technol.* 2015;25(4):317–327.
- [9] Zhang Y, Dong Z, Liu A, et al. Magnetic resonance brain image classification via stationary wavelet transform and generalized eigenvalue proximal support vector machine. *J Med Imaging Health Inform.* 2015;5(7):1395–1403.
- [10] Li Q, Cai W, Wang X, et al. Medical image classification with convolutional neural network. Proceedings of 13th international conference on control automation robotics & vision (ICARCV), IEEE, 844–848, Singapore, December 2014.
- [11] Bar Y, Diamant I, Wolf L, et al. Deep learning with non-medical training used for chest pathology identification. Proceedings of SPIE medical imaging, international society for optics and photonics, 94140V, Orlando, FL, USA, February 2015.
- [12] Shin H-C, Roth HR, Gao M, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans Med Imaging.* 2016;35(5):1285–1298.
- [13] Barata C, Ruela M, Francisco M, et al. Two systems for the detection of melanomas in dermoscopy images using texture and color features. *IEEE Syst J.* 2014;8(3):965–979.
- [14] Iyatomi H, Oka H, Celebi ME, et al. An improved internet-based melanoma screening system with dermatologist-like tumor area extraction algorithm. *Comput Med Imaging Graph.* 2008;32(7):566–579.
- [15] Stoecker WV, Wronkiewicz M, Chowdhury R, et al. Detection of granularity in dermoscopy images of malignant melanoma using color and texture features. *Comput Med Imaging Graph.* 2011;35(2):144–147.
- [16] Ahn E, Kumar A, Kim J, et al. X-ray image classification using domain transferred convolutional neural networks and local sparse spatial pyramid. Proceedings of 2016, IEEE international symposium on biomedical imaging (ISBI), 855–858, IEEE, Prague, Czech Republic, April 2016.
- [17] Ouyang W, Luo P, Zeng X, et al. “DeepID-net: multi-stage and deformable deep convolutional neural networks for object detection,” <http://arxiv.org/abs/1409.3505>.
- [18] Fernando B, Fromont E, Muselet D, et al. Discriminative feature fusion for image classification). Proceedings of 2012 IEEE conference on computer vision and pattern recognition (CVPR), 3434–3441, IEEE, Providence, RI, USA, June 2012.
- [19] Ramirez J, Gorritz JM, Segovia F, et al. Computer aided diagnosis system for the Alzheimer’s disease based on partial least squares and random forest spect image classification. *Neurosci Lett.* 2010;472(2):99–103.
- [20] Zeng R, Wu J, Shao Z, et al. Color image classification via quaternion principal component analysis network. *Neurocomputing.* 2016;216:416–428.
- [21] Chan T, Jia K, Gao S, et al. PCANet: A simple deep learning baseline for image classification? *IEEE Trans Image Process.* 2015;24(12):5017–5032.
- [22] Rakotomamonjy A, Petitjean C, Salaun M, et al. Scattering features for lung cancer detection in fibered confocal fluorescence microscopy images. *Artif Intell Med.* 2014;61(2):105–118.
- [23] Bruna J, Mallat S. Invariant scattering convolution networks. *IEEE Trans Pattern Anal Mach Intell.* 2013;35(8):1872–1886.
- [24] Cruzroa A, Basavanahally A, Gonzalez FA, et al. Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks. Proceedings of SPIE, 904103, Houston, TX, USA, May 2014.
- [25] Chiang YY, Knoblock CA. A general approach for extracting road vector data from raster maps. *Int J Doc Anal Recognit.* 2013;16(1):55–81. doi:10.1007/s10032-011-0177-1.
- [26] Bonnici A, Camilleri K. A circle-based vectorization algorithm for drawings with shadows. Proceedings of the international symposium on sketch-based interfaces and modeling, 69–77, ACM, New York, NY, USA, 2013. doi:10.1145/2487381.2487386.
- [27] Xia T, Liao B, Yu Y. Patch-based image vectorization with automatic curvilinear feature alignment. ACM SIGGRAPH Asia 2009 papers, SIGGRAPH Asia ‘09, 115:1–115:10, ACM, New York, NY, USA, 2009. doi:10.1145/1661412.1618461.
- [28] Bo P, Luo G, Wang K. A graph-based method for fitting planar b-spline curves with intersections. *J Comput Des Eng.* 2016;3(1):14–23. <https://doi.org/10.1016/j.jcde.2015.05.001>.