



# Automated Medical Diagnosis System Based on Multi-modality Image Fusion and Deep Learning

Abeer D. Algarni<sup>1</sup>

Published online: 1 November 2019  
© Springer Science+Business Media, LLC, part of Springer Nature 2019

## Abstract

Multi-modality medical image fusion aims at integrating information from medical images with different modalities to aid in the diagnosis process. Most research work in this area ends with the fusion stage only. This paper, on the contrary, tries to present a complete diagnosis system based on multi-modality image fusion. This system works on MR and CT images. It begins with the registration step using Scale-Invariant Feature Transform registration algorithm. After that, histogram matching is performed to allow accurate fusion of the medical images. Two methods of the fusion are utilized and compared, wavelet and curvelet fusion. An interpolation stage is included to enhance the resolution of the obtained image after fusion. Finally, a deep learning approach is adopted for classification of images as normal or abnormal. Simulation results reveal good success of the proposed automated diagnosis system based on the fusion and interpolation results.

**Keywords** Image fusion · MR · CT · Registration · Interpolation · Classification · Deep learning

## 1 Introduction

Medical imaging has become a very powerful and efficient tool that can be used in the diagnosis of several diseases. There are several technologies to extract medical images such as X-ray imaging, mammography, Computed Tomography (CT), Computerized Axial Tomography (CAT), Magnetic Resonance Imaging (MRI), ultrasonic imaging, Positron Emission Tomography (PET), and confocal microscopy imaging. These imaging technologies have different principles for generating images, and hence different artifacts and limitations [1, 2]. In some scenarios, there is a need to integrate information from different medical images obtained with different modalities for the same object to help in a better diagnosis. The most popular example of this is the fusion of MR and CT images. To achieve a successful fusion process, there is a need first to understand the nature of both modalities of images.

---

✉ Abeer D. Algarni  
adalqarni@pnu.edu.sa

<sup>1</sup> Department of Information Technology, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, Riyadh, Saudi Arabia

CT imaging is based on X-rays by collecting several angular-projection X-ray images. Cross section slices can be generated from these images using different reconstruction algorithms. Some artifacts such as ringing, high attenuation around edges, ghosting, and stitching may appear in CT images. These artifacts need to be removed with efficient image processing algorithms [1, 2]. On the other hand, the idea of MRI differs from that of X-ray imaging. Here, no radiation is used. This technique just depends on a strong magnetic field, which causes hydrogen atoms to emit radio-wave pulses that can be used for imaging. Concentration of hydrogen atoms is reflected in the obtained MRI images, and hence tissues can be discriminated based on the amount of water in each of them. Several artifacts may appear in MRI images such as zebra stripes, inhomogeneity, shading, and aliasing [1, 2].

Multi-modality image fusion is used to merge information from multiple images obtained with different imaging techniques. The most common types of images that are incorporated in the fusion process include MR and CT images. These images need first to be registered for accurate fusion results. Different approaches have been presented for this objective. Ali et al. presented an approach for MR and CT image fusion based on both wavelet transform and curvelet transform [3, 4]. Unfortunately, this approach did not consider the histogram matching issue, and hence quality of the obtained fusion images is not good. Moreover, Ali et al. developed their work for resolution enhancement of medical images by incorporating inverse interpolation techniques such as maximum entropy interpolation, LMMSE interpolation, and regularized interpolation [5]. This work presented images of better resolution, but unfortunately did not consider the histogram matching issue. The main shortcoming of this work is that it was not extended to the diagnosis process. It stopped at the fusion and interpolation stages only. To be more feasible, there is a need to go to the further classification stage.

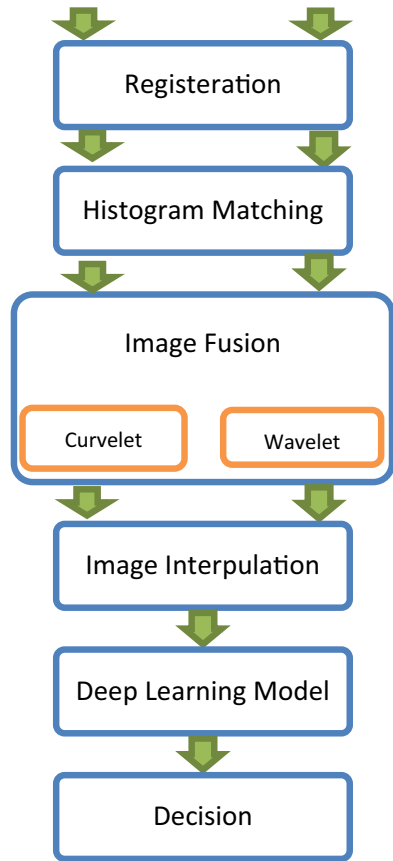
Some other attempts have been presented in the literature to combine the process of multi-modality image fusion with optimization to enhance the quality of the obtained fusion results subject to certain constraints [6, 7]. These attempts considered optimization techniques like the modified central force optimization in maximizing the visual quality of the obtained results. The attempts in [6, 7] considered the histogram matching issue and gave good fusion results. Unfortunately, these techniques did not consider the resolution enhancement scenario. They did not also consider the automated diagnosis process needed after obtaining the fusion results.

The work in this paper differs from the work presented by Ali et al. in that it considers the histogram matching as a very major stage to match the intensity levels of the CT image to those of the MR image for better fusion results. Moreover, the proposed work in this paper presents a full diagnosis system by introducing a classification stage for the fusion results based on a suggested deep learning model. Simulation results presented in the paper prove the feasibility and success of the proposed automated medical diagnosis system to differentiate between fusion results with tumors and fusion results without tumors.

## 2 Proposed Automated Diagnosis System

The proposed automated diagnosis system consists of five stages as shown in Fig. 1. These stages are registration, histogram matching, image fusion, image interpolation, and deep learning for classification. The first stage is the registration stage. This stage is performed to calibrate the coordinates of both MR and CT images to allow an accurate fusion process. The registration process can be performed using SIFT algorithm [8–11]. The SIFT algorithm is used

Fig. 1 The proposed approach



to extract feature points across scales and represent these feature points with feature vectors containing angle histogram bins. The registration is performed by rotating one of the images with fixed step sizes and comparing the feature vectors between the reference image and each rotated image [8–11].

### 3 Registration of MR and CT Images

To extract the feature vector from an image, the first step is to convolve the image at hand with multiple Gaussian kernels with different scales [8–11]:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (1)$$

where  $\sigma$ , and  $G(x, y, \sigma)$  are the scale and the Gaussian kernel with scale  $\sigma$ , and  $I(x, y)$  is the original image. The Gaussian kernel is defined as follows [8–11]:

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad (2)$$

Laplacian operator is applied to the different scales. In a simplified manner, the Difference of Gaussina (DoG) is used as an alternative strategy as illustrated in Fig. 2 for the determination of the keypoints. A comparison process is performed between pixels at similar coordinates to select the feature points across scales. The extrema points across scales are selected for further feature extraction processes. The DoG is a good alternative to the LoG as the Laplacian is eliminated for simplification [8–11].

After keypoint selection, both magnitude and phase of the gradient at this keypoint are estimated as follows [8–11]:

$$m(x, y) = \sqrt{(L(x + 1, y) - L(x - 1, y))^2 + (L(x, y + 1) - L(x, y - 1))^2} \quad (3)$$

$$\theta(x, y) = \tan^{-1} \left( \frac{L(x, y + 1) - L(x, y - 1)}{L(x + 1, y) - L(x - 1, y)} \right) \quad (4)$$

where  $L(x, y)$  is the gradient at  $(x, y)$ . A  $16 \times 16$  window around each keypoint is selected and divided into four sub-windows as shown in Fig. 3. An angle histogram of gradients is estimated in each sub-window with four bins, and each bin represents  $45^\circ$ . This leads to a feature vector of 128 points representing each feature point.

The registration process is performed by taking the MR image as a reference image and extracting the feature points and feature vectors from that image. The feature points and feature vectors are also extracted from the CT image and matched to those of the MR image. The matching is based on the minimum distance criterion. To accommodate for rotation effects, a group of rotations of the CT image with step angles are performed and followed by a matching process after each rotation. The feature comparison is performed for each rotation angle and the decision of registration is taken based on the minimum distance criterion.

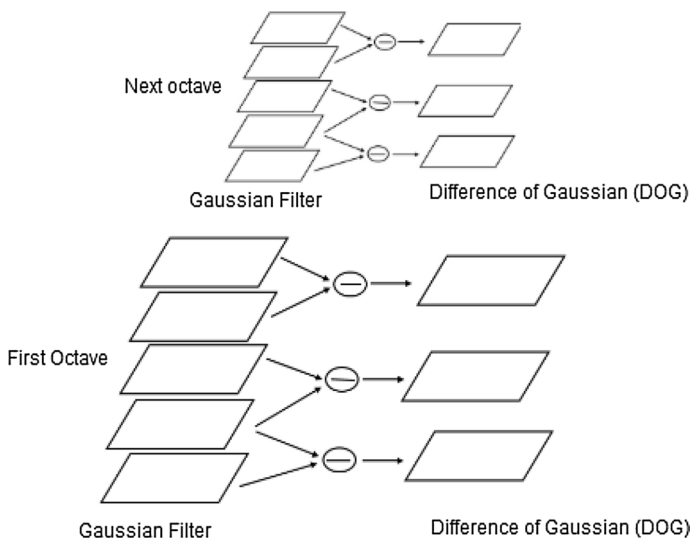


Fig. 2 Implementation of the DoG

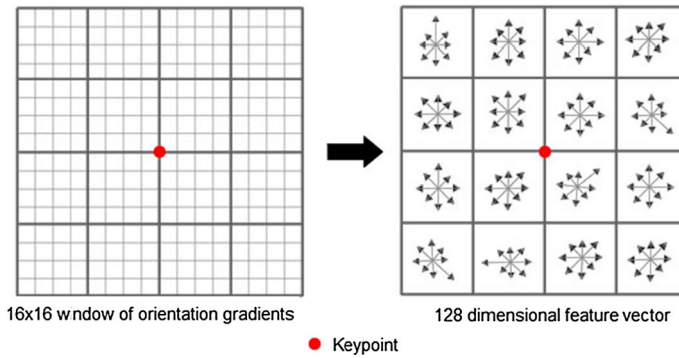


Fig. 3 Keypoint description

## 4 Histogram Matching

Histogram matching is a process that aims to adapt the histogram of an image to that of another one for the objective of enhancement of a poor-visual-quality image as in the case of infrared image enhancement [12, 13], or the objective of image fusion. It has become a must in multi-modality image fusion, because the intensity levels of the multi-modality images differ so much. In the scenario at hand of fusion of MR and CT images, the CT image is taken as the reference image  $f(k,l)$  as it has a larger dynamic range, and the MR image intensities  $g(m,n)$  are adjusted based on this reference image. The steps of the histogram matching algorithm adopted in this paper is similar to that adopted by Ashiba et al. for infrared image enhancement and its steps are summarized below [12].

1. Estimate the mean of the MR image.

$$\hat{g} = \left( \frac{1}{MN} \right) \sum_{m=1}^M \sum_{n=1}^N g(m,n) \quad (5)$$

where  $M$  and  $N$  are the dimensions of the MR image.

2. Estimate the mean of the CT image.

$$\hat{f} = \left( \frac{1}{KL} \right) \sum_{k=1}^K \sum_{l=1}^L f(k,l) \quad (6)$$

where  $K$  and  $L$  are the dimensions of the CT image.

3. Estimate the standard deviation of the MR image.

$$\sigma_1 = \sqrt{\frac{1}{MN} \sum_{m=1}^M \sum_{n=1}^N (g(m,n) - \hat{g})^2} \quad (7)$$

4. Estimate the standard deviation of the CT image.

$$\sigma_2 = \sqrt{\frac{1}{KL} \sum_{k=1}^K \sum_{l=1}^L (f(k, l) - \hat{f})^2} \quad (8)$$

5. Estimate the correction factor.

$$C = \frac{\sigma_2}{\sigma_1} \quad (9)$$

6. Estimate the corrected mean of the MR image.

$$f_c = \hat{f} - C \cdot \hat{g} \quad (10)$$

7. Estimate the histogram matched MR image.

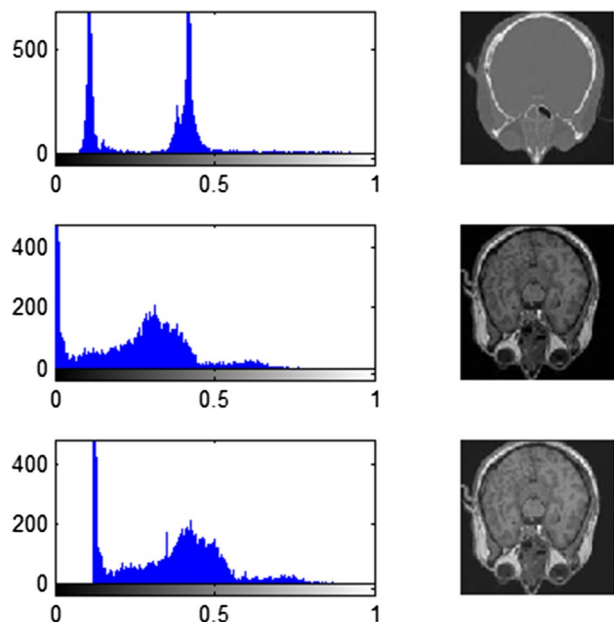
$$g_h(m, n) = f_c + C \cdot g(m, n) \quad (11)$$

An example of the adjustment of an MR image histogram to that of a CT image histogram is shown in Fig. 4. It is clear that after histogram matching, the intensities of the MR image become brighter to accommodate for the intensities of the CT image.

## 5 Fusion of MR and CT Images

Fusion of MR and CT images is performed in this paper with either wavelet or curvelet fusion. They are implemented and compared.

**Fig. 4** The results of histogram matching of an MR image to a CT image



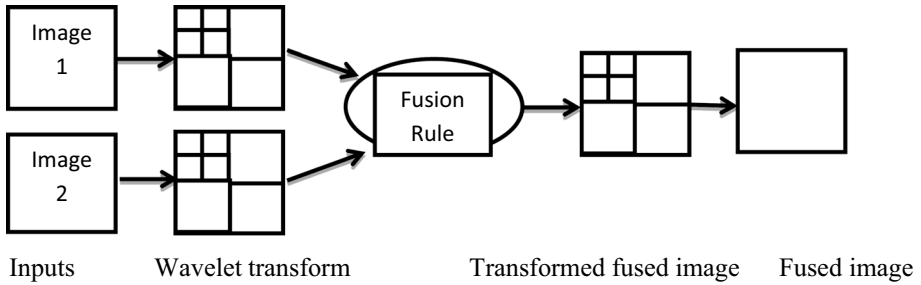


Fig. 5 Wavelet fusion of MR and CT images

### 5.1 Wavelet Fusion of MR and CT Images

Wavelet transform is utilized as a tool for the fusion of the MR and CT images after the histogram matching step as shown in Fig. 5. The wavelet transform is applied on both images and the fusion rule is applied on the wavelet coefficients as follows [3–5]:

$$I_f(x, y) = IDWT[\emptyset\{DWT(I_1(n_1, n_2)) \cdot DWT(I_2(n_1, n_2))\}] \tag{12}$$

where DWT refers to the wavelet transform in two dimensions, IDWT refers to the inverse wavelet transform, and  $\emptyset$  is the fusion rule.

The fusion rule adopted in this paper is the maximum fusion rule, because with this rule, it is guaranteed that any sort of blurring in any of the images to be fused is eliminated. The wavelet transform is reported to give good fusion results, but it has a problem with the fusion of curved edges. So, the curvelet transform is a good candidate for this task.

### 5.2 Curvelet Fusion of MR and CT Images

The curvelet fusion of MR and CT images is based on the curvelet transform of both images. The curvelet transform is implemented on overlapping tiles of the images to approximate curved lines as short straight lines after applying the Additive Wavelet Transform (AWT) as shown in Fig. 6. Ridgelet transform is applied on each tile, and then maximum fusion rule is implemented in the fusion process [3–5].

First of all, the additive wavelet transform is applied on both MR and CT images. The steps of the AWT are shown in Fig. 7. It is implemented through cascaded filtering stages and difference operations to get the detail planes and finally an approximation plane.

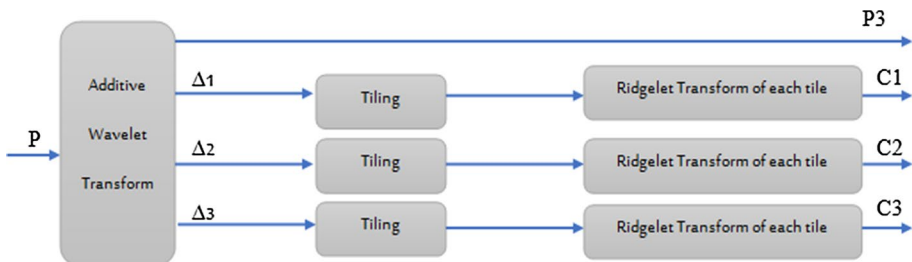


Fig. 6 Curvelet transform of an image

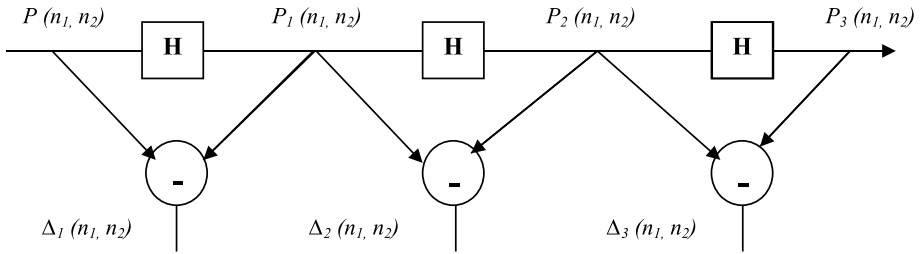


Fig. 7 Steps of the AWT of an image

Given an image  $P$ , it is possible to construct the sequence of approximations as follows [3–5]:

$$f_1(P) = P_1, f_2(P_1) = P_2, f_3(P_2) = P_3, \dots, f_n(P_{n-1}) = P_n, \tag{13}$$

where  $n$  is an integer. In constructing the above sequence, successive convolutions with the kernel  $H$  are performed,

$$H = \frac{1}{256} \begin{pmatrix} 1 & 4 & 6 & 4 & 1 \\ 4 & 16 & 24 & 16 & 4 \\ 6 & 24 & 36 & 24 & 6 \\ 4 & 16 & 24 & 16 & 4 \\ 1 & 4 & 6 & 4 & 1 \end{pmatrix} \tag{14}$$

The wavelet planes are calculated as the differences between two consecutive approximations  $P_{l-1}$  and  $P_l$ , i.e. [3–5].

$$\Delta_l = P_{l-1} - P_l \tag{15}$$

Therefore, the curvelet reconstruction formula is given by [3–5]:

$$P = \sum_{l=1}^{n-1} \Delta_l + P_n \tag{16}$$

This step is performed after the fusion of the corresponding detail components of the MR and CT tiles, while keeping the approximation of the MR image as it is rich in details.

The ridgelet transform is applied to the tiles of the images prior to the fusion process. Assume a tile  $f(x_1, x_2)$ , the ridgelet transform of that tile is given by [3–5]:

$$R_f(p, q, \theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \varphi_{p,q,\theta}(x_1, x_2) f(x_1, x_2) dx_1 dx_2 \tag{17}$$

where  $\varphi_{p,q,\theta}(x_1, x_2) = a^{-\frac{1}{2}} \varphi\left(\frac{x_1 \cos \theta + x_2 \sin \theta - b}{a}\right)$  is the ridgelet basis function,  $a > 0$  and  $\theta \in [0, 2\pi]$  [3–5].

The ridgelet transform is closely related to the Radon transform defined as,

$$R_f(\theta, t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, x_2) \delta(x_1 \cos \theta + x_2 \sin \theta - t) dx_1 dx_2 \tag{18}$$



This relation is given by [3–5]:

$$R_f(p, q, \theta) = \int_{-\infty}^{\infty} R_f(\theta, t) a^{-\frac{1}{2}} \left( \frac{t-b}{a} \right) dt \tag{19}$$

The main advantage of this relation is that the projection slice theorem of radon transform is combined with ridgelet transform to transform line singularities to point singularities.

### 6 Image Interpolation

After obtaining the fused image, it is required to increase the resolution of this image for better diagnosis. To achieve this objective, several interpolation techniques can be investigated. Prior to going to the interpolation problem, there is a need to understand the relation between the available fusion result and the required high-resolution image. This relation can be defined as follows [14–16]:

$$\mathbf{g} = \mathbf{D}\mathbf{f} + \mathbf{v} \tag{20}$$

where  $\mathbf{g}$ ,  $\mathbf{f}$ ,  $\mathbf{D}$ , and  $\mathbf{v}$  are the obtained fusion result in lexicographic format, the required high-resolution image in lexicographic format, the decimation matrix, and the noise vector, respectively.

The matrix  $\mathbf{D}$  is defined as:

$$\mathbf{D} = \mathbf{D}_1 \otimes \mathbf{D}_1 \tag{21}$$

where  $\otimes$  represents the Kronecker product, and the  $M \times N$  matrix  $\mathbf{D}_1$  represents the 1-D filtering and down-sampling by a factor  $R$ . For  $N = 2M$ ,  $\mathbf{D}_1$  is given by [14–16]:

$$\mathbf{D}_1 = \frac{1}{2} \begin{bmatrix} 1 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & 1 \end{bmatrix} \tag{22}$$

The model of filtering and down-sampling is illustrated in Fig. 8. The objective of the image interpolation process is the estimation of the vector  $\mathbf{f}$  given the vector  $\mathbf{g}$ . Several algorithms can be used to achieve this objective.

#### 6.1 Cubic Spline Image Interpolation

The signal synthesis approach to solve the interpolation problem is to use the following equation based on Fig. 9 [14–16]:

$$\hat{f}(x) = \sum_{k=-\infty}^{\infty} c(x_k) \beta(x - x_k) \tag{23}$$

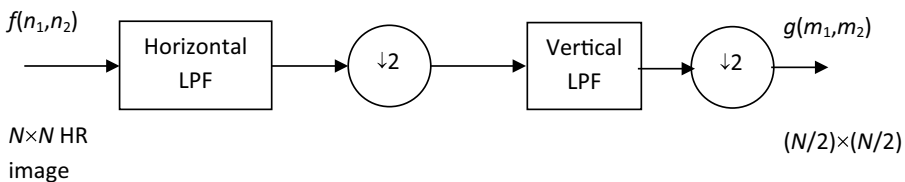
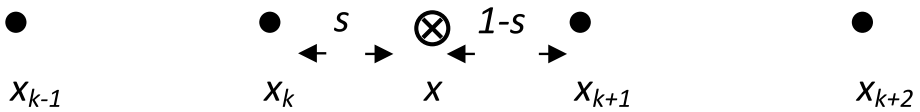


Fig. 8 Down-sampling process from the  $N \times N$  HR image to the  $(N/2) \times (N/2)$  LR image



**Fig. 9** 1-D signal interpolation. The pixel at position  $x$  is estimated using its neighborhood pixels and the distance  $s$

A good approximation to solve the interpolation equation is to use a spline of order  $n$  as follows [14–16]:

$$\hat{f}(x) = \sum_{k \in Z} c(x_k) \beta^n(x - x_k) \tag{24}$$

where  $Z$  is a finite neighborhood around  $x$ .

The B-spline function  $\beta^n(x)$  is obtained by  $(n + 1)$  fold convolutions of  $\beta^0$  given by [14–16]:

$$\beta^0(x) = \begin{cases} 1 & -\frac{1}{2} < x < \frac{1}{2} \\ \frac{1}{2} & |x| = \frac{1}{2} \\ 0 & \text{Otherwise} \end{cases} \tag{25}$$

The most popular spline interpolation basis function is the cubic basis function defined as [14–16]:

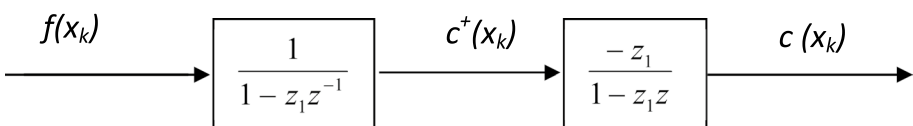
$$\beta^3(x) = \begin{cases} \frac{2}{3} - |x|^2 + \frac{|x|^3}{2} & 0 \leq |x| < 1 \\ \frac{(2-|x|)^3}{6} & 1 \leq |x| < 2 \\ 0 & 2 \leq |x| \end{cases} \tag{26}$$

Solving Eq. (20) with the aid of Fig. 10, we get the following closed-form expression for the cubic spline interpolation formula.

$$\begin{aligned} \hat{f}(x) = & c(x_{k-1})[(3 + s)^3 - 4(2 + s)^3 + 6(1 + s)^3 - 4s^3]/6 \\ & + c(x_k)[(2 + s)^3 - 4(1 + s)^3 + 6s^3]/6 + c(x_{k+1})[(1 + s)^3 - 4s^3]/6 + c(x_{k+2})s^3/6 \end{aligned} \tag{27}$$

The cubic spline basis function is non-interpolating, and thus the coefficients in Eq. (27) need to be estimated prior to the interpolation process. The estimation process can be implemented using a digital filtering approach [17–19].

To begin the digital filtering algorithm required for the determination of the spline coefficients as shown in Fig. 11, we define the B-spline kernel  $b_m^n$ , which is obtained by sampling the B-spline basis function of degree  $n$  expanded by a factor  $m$  [17–19].



**Fig. 10** Estimation of the cubic spline interpolation coefficients

$$b_m^n(x_k) = \beta^n(x/m)|_{x=x_k} \tag{28}$$

The z-transform of the B-spline kernel is given by:

$$B_m^n(z) = \sum_{x_k \in \mathbb{Z}} b_m^n(x_k) z^{-k} \tag{29}$$

The main objective is to estimate the coefficients  $c(x_k)$  to achieve a perfect fit at the integers, which means that [17–19]:

$$\sum_{x_l \in \mathbb{Z}} c(x_l) \beta^n(x - x_l)|_{x=x_k} = f(x_k) \tag{30}$$

Using convolution notation [17–19]:

$$f(x_k) = (b_1^n * c)(x_k) \tag{31}$$

Hence, the coefficients are given by [17–19]:

$$c(x_k) = (b_1^n)^{-1} * f(x_k) \tag{32}$$

Since  $b_1^n$  is a symmetric FIR filter, the direct B-Spline filter  $(b_1^n)^{-1}$  is an all-pole filter that can be implemented using a cascade of first order causal and non-causal recursive filters [17–19].

For the case of cubic spline interpolation, we have [17–19]:

$$B_1^3(z) = \frac{z + 4 + z^{-1}}{6} \tag{33}$$

This gives:

$$[B_1^3(z)]^{-1} = 6 \left[ \frac{1}{1 - z_1 z^{-1}} \right] \left[ \frac{-z_1}{1 - z_1 z} \right] \tag{34}$$

where  $z_1 = -2 + \sqrt{3}$ .

The above equation can be implemented with two cascaded stages as illustrated in Fig. 10.

$$c^+(x_k) = f(x_k) + z_1 c^+(x_{k-1}) \quad (k = 1, \dots, N - 1) \tag{35}$$

$$c(x_k) = z_1 (c(x_{k+1}) - c^+(x_k)), \quad (k = N - 2, \dots, 0) \tag{36}$$

$N$  is the length of the data sequence used.

The initial conditions for the two stages can be calculated as follows [17–19]:

$$c^+(0) = \sum_{k=0}^{N-1} f(x_k) z_1^k \tag{37}$$

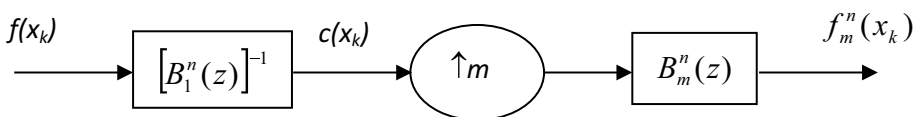


Fig. 11 Block diagram of B-spline image interpolation

$$c(x_{N-1}) = \frac{z_1}{(1 - z_1^2)} (c^+(x_{N-1}) + z_1 c^+(x_{N-2})) \tag{38}$$

The block diagram of the image interpolation by an up-sampling factor  $m$  is illustrated in Fig. 11 in an all-digital implementation. We denote the interpolated image as  $f_m^n(x_k)$ , which can be obtained using the following equation [17–19]:

$$f_m^n(x_k) = \sum_{l=-\infty}^{\infty} c(x_l) b_m^n(x_k - mx_l) \tag{39}$$

The digital filter  $B_m^n(z)$  is a symmetric FIR filter. The expression for  $B_m^n(z)$  is given as follows [17–19]:

$$B_m^n(z) = \frac{z^\alpha}{m^n} \left( \frac{1 - z^{-m}}{1 - z^{-1}} \right)^{n+1} B_1^n(z) \tag{40}$$

where  $\alpha = (m - 1)(n + 1)/2$ .

### 6.2 Inverse Interpolation

The treatment of the interpolation as an inverse problem depends on the inversion of Eq. (20). Several models can be used for this objective such as the Linear Minimum Mean Square Error (LMMSE), the maximum entropy, and the regularized models. In order to set the LMMSE solution, we need to solve [14–16]:

$$\min_{\hat{\mathbf{f}}} E[\mathbf{e}'\mathbf{e}] = E[\text{Tr}(\mathbf{e}\mathbf{e}')] \tag{41}$$

with

$$\mathbf{e} = \mathbf{f} - \hat{\mathbf{f}} \tag{42}$$

where  $\hat{\mathbf{f}}$  is the estimate of the required HR image.

If we assume that:

$$\hat{\mathbf{f}} = \mathbf{T}\mathbf{g} \tag{43}$$

The operator  $\mathbf{T}$  is derived subject to minimizing Eq. (41) to give [14–16]:

$$\begin{aligned} \min_{\hat{\mathbf{f}}} E[\text{Tr}(\mathbf{e}\mathbf{e}')] &= E[\text{Tr}\{(\mathbf{f} - \mathbf{T}\mathbf{g})(\mathbf{f} - \mathbf{T}\mathbf{g})'\}] \\ &= E[\text{Tr}\{\mathbf{f}\mathbf{f}' - \mathbf{T}(\mathbf{D}\mathbf{f}\mathbf{f}' + \mathbf{v}\mathbf{f}') - (\mathbf{f}\mathbf{f}'\mathbf{D}' + \mathbf{f}\mathbf{v}')\mathbf{T}' + \mathbf{T}(\mathbf{D}\mathbf{f}\mathbf{f}'\mathbf{D}' + \mathbf{v}\mathbf{f}'\mathbf{D}' + \mathbf{D}\mathbf{f}\mathbf{v}' + \mathbf{v}\mathbf{v}')\mathbf{T}'\}] \end{aligned} \tag{44}$$

We have  $\text{Tr}(\mathbf{B}) = \text{Tr}(\mathbf{B}')$ . Since the trace is linear, it can be interchanged with the expectation operator.

Equation (44) can be simplified with the assumptions:

$$E[\mathbf{f}\mathbf{v}'] = E[\mathbf{v}'\mathbf{f}] = [\mathbf{0}] \tag{45}$$

The autocorrelation matrices for the image and noise can be defined as:

$$E[\mathbf{f}\mathbf{f}^t] = \mathbf{R}_f \quad (46)$$

and

$$E[\mathbf{v}\mathbf{v}^t] = \mathbf{R}_v \quad (47)$$

The matrix  $\mathbf{R}_v$  is a diagonal matrix whose main diagonal elements are equal to the noise variance in the noisy LR image.

This leads to the minimization problem:

$$\min_{\hat{\mathbf{f}}} E[\text{Tr}(\mathbf{e}\mathbf{e}^t)] = \text{Tr}\{\mathbf{R}_f - 2\mathbf{T}\mathbf{D}\mathbf{R}_f + \mathbf{T}\mathbf{D}\mathbf{R}_f\mathbf{D}^t\mathbf{T}^t + \mathbf{T}\mathbf{R}_v\mathbf{T}^t\} \quad (48)$$

Differentiating Eq. (48) with respect to  $\mathbf{T}$  and setting the result equal to zero, the operator  $\mathbf{T}$  is given by:

$$\mathbf{T} = \mathbf{R}_f\mathbf{D}'(\mathbf{D}\mathbf{R}_f\mathbf{D}' + \mathbf{R}_v)^{-1} \quad (49)$$

Thus, the LMMSE estimate of the HR image will be given by [14–16]:

$$\hat{\mathbf{f}} = \mathbf{R}_f\mathbf{D}'(\mathbf{D}\mathbf{R}_f\mathbf{D}' + \mathbf{R}_v)^{-1}\mathbf{g} \quad (50)$$

To solve the problem of estimating the autocorrelation of the HR image, the matrix  $\mathbf{R}_f$  in Eq. (50) can be written in the form [14–16]:

$$\mathbf{R}_f = \begin{bmatrix} \mathbf{R}_{0,0} & \mathbf{R}_{0,1} & \cdots & \mathbf{R}_{0,N-1} \\ \mathbf{R}_{1,0} & \mathbf{R}_{1,1} & \cdots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{R}_{N-1,0} & \cdots & \cdots & \mathbf{R}_{N-1,N-1} \end{bmatrix} \quad (51)$$

where

$$\mathbf{R}_{i,j} = E[\mathbf{f}_i\mathbf{f}_j^t] \quad (52)$$

$\mathbf{f}_i$  and  $\mathbf{f}_j$  are the  $i$ th and  $j$ th column partitions of the lexicographically-ordered vector  $\mathbf{f}$ . Usually, pixels in an image have no predictable correlation beyond a correlation distance  $d$ . If we assume that  $d=0$ , then the matrix  $\mathbf{R}_f$  can be dealt with as a diagonal matrix in the form [14–16]:

$$\mathbf{R}_f = \begin{bmatrix} \mathbf{R}_{0,0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{1,1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{R}_{N-1,N-1} \end{bmatrix} \quad (53)$$

If the pixels of each column are uncorrelated except for each pixel with itself, each matrix  $\mathbf{R}_{i,i}$  can be dealt with as a diagonal matrix for  $i=0,1,\dots,N-1$  as follows [14–16]:

$$\mathbf{R}_{i,i} = \begin{bmatrix} R_f(i,0) & 0 & \cdots & 0 \\ 0 & R_f(i,1) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & R_f(i,N-1) \end{bmatrix} \quad (54)$$

The diagonal elements of the matrix  $\mathbf{R}_{i,i}$  can be estimated from a polynomial-based interpolated version of the available LR image.

For an image  $f'(n_1, n_2)$ , the autocorrelation at the spatial location  $(n_1, n_2)$  can be estimated from the following relation [14–16]:

$$R_f(n_1, n_2) \cong \frac{1}{w^2} \sum_{k=1}^w \sum_{l=1}^w f'(k, l) f'(n_1 + k, n_2 + l) \tag{55}$$

where  $R_f(n_1, n_2)$  is the autocorrelation at spatial position  $(n_1, n_2)$  and  $w$  is an arbitrary window length for the auto-correlation estimation.

The image  $f'(n_1, n_2)$  may be taken as the cubic spline interpolated image. Thus, the matrix  $\mathbf{R}_f$  can be approximated by a diagonal sparse matrix.

Another inverse solution is to adopt the maximum entropy concept to obtain the HR image. The required HR image is assumed to be treated as light quanta associated with each pixel value. Thus, the entropy of the required HR image is defined as follows [14–16]:

$$H_e = - \sum_{i=1}^{N^2} f_i \log_2(f_i) \tag{56}$$

where  $H_e$  is the entropy and  $f_i$  is the sampled signal.

This equation can be written in the vector form as follows [14–16]:

$$H_e = -\mathbf{f}' \log_2(\mathbf{f}) \tag{57}$$

For image interpolation, to maximize the entropy subject to the constraint that  $\|\mathbf{g} - \mathbf{D}\mathbf{f}\|^2 = \|\mathbf{v}\|^2$ , the following cost function must be minimized [14–16]:

$$\Psi(\mathbf{f}) = \mathbf{f}' \log_2(\mathbf{f}) - \lambda [\|\mathbf{g} - \mathbf{D}\mathbf{f}\|^2 - \|\mathbf{v}\|^2] \tag{58}$$

where  $\lambda$  is a Lagrangian multiplier.

Differentiating both sides of the above equation with respect to  $\mathbf{f}$  and equating the result to zero:

$$\frac{\partial \Psi(\mathbf{f})}{\partial \mathbf{f}} = 0 = \frac{1}{\ln(2)} \left\{ \mathbf{1} + \ln(\hat{\mathbf{f}}) \right\} + \lambda [2\mathbf{D}'(\mathbf{g} - \mathbf{D}\hat{\mathbf{f}})] \tag{59}$$

Solving for the estimated HR image:

$$\ln(\hat{\mathbf{f}}) = -\mathbf{1} - \lambda \ln(2) [2\mathbf{D}'(\mathbf{g} - \mathbf{D}\hat{\mathbf{f}})] \tag{60}$$

Thus:

$$\hat{\mathbf{f}} = \exp \left[ -\mathbf{1} - \lambda \ln(2) [2\mathbf{D}'(\mathbf{g} - \mathbf{D}\hat{\mathbf{f}})] \right] \tag{61}$$

Expanding the above equation using Taylor expansion and neglecting all but the first two terms, and since  $\mathbf{g} - \mathbf{D}\hat{\mathbf{f}}$  must be a small quantity, the following form is obtained:

$$\hat{\mathbf{f}} \cong -\lambda \ln(2) [2\mathbf{D}'(\mathbf{g} - \mathbf{D}\hat{\mathbf{f}})] \tag{62}$$

Solving for  $\hat{\mathbf{f}}$  leads to [14–16]:

$$\hat{\mathbf{f}} \cong (\mathbf{D}'\mathbf{D} + \eta\mathbf{I})^{-1} \mathbf{D}'\mathbf{g} \tag{63}$$

where  $\eta = -1/(2\lambda \ln(2))$ .

The operation  $\mathbf{D}'\mathbf{D}$  stands for decimation followed by interpolation. Thus, if  $\mathbf{D}$  decimates by a factor  $r$ , applying  $\mathbf{D}'\mathbf{D}$  causes all positions  $(1+n_1r, 1+n_2r)$  for integer  $(n_1, n_2)$  to stay unchanged, whereas the remaining pixels are replaced by zeros. Thus,  $\mathbf{D}'\mathbf{D}$  stands for a masking operation, which is represented by a diagonal matrix. The effect of the term  $\eta\mathbf{I}$  is to remove the ill-posedness nature of the inverse problem. The problem now is reduced to an inversion of a diagonal matrix.

Another solution is based on the regularization theory by modifying the cost function to be [14–16]:

$$\Psi(\hat{\mathbf{f}}) = \|\mathbf{g} - \mathbf{D}\hat{\mathbf{f}}\|^2 + \lambda\|\mathbf{Q}\hat{\mathbf{f}}\|^2 \tag{64}$$

where  $\mathbf{Q}$  is the 2-D regularization operator and  $\lambda$  is the regularization parameter.

This minimization is performed by taking the derivative of the cost function yielding:

$$\frac{\partial\Psi(\hat{\mathbf{f}})}{\partial\hat{\mathbf{f}}} = \mathbf{0} = 2\mathbf{D}'(\mathbf{g} - \mathbf{D}\hat{\mathbf{f}}) - 2\lambda\mathbf{Q}'\mathbf{Q}\hat{\mathbf{f}} \tag{65}$$

Solving for  $\hat{\mathbf{f}}$  that provides the minimum of the cost function yields [14–16]:

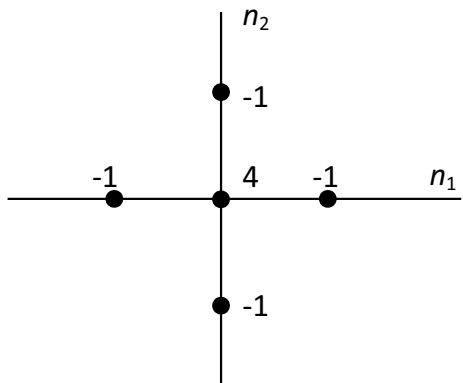
$$\hat{\mathbf{f}} = (\mathbf{D}'\mathbf{D} + \lambda\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{D}'\mathbf{g} \tag{66}$$

The regularization term in Eq. (66) allows matrix inversion even in the presence of singularities. This equation can be implemented on a block-by-block basis using overlapping blocks. The operator  $\mathbf{Q}$  can be represented with the Laplacian illustrated in Fig. 12.

## 7 Deep Learning Classification

The deep learning classification is the last task in the proposed automated diagnosis system for tumor detection. Deep learning is a type for machine learning in which a model learns to perform classification tasks directly from images. Features are learnt automatically from the obtained HR fusion results by a convolutional neural network (CNN) and presented to a classifier. The CNN is a neural network for image recognition. It includes the feature extractor in the training process [20–23]. It consists of convolutional layers followed by activation functions, pooling layers, dropout layers and fully-connected layers.

Fig. 12 The 2-D Laplacian operator



The inclusion of a dropout layer adds a regularization technique for reducing over fitting [20–23].

## 7.1 Convolutional Layer (CNV)

The convolutional layer contains filters that are used to perform a two-dimensional (2D) convolution on the input image. The resulting features of the convolutional layer vary depending on which convolution filter is used. This concept is very well suited to the HR fusion results as it is able to capture any slight change in the image local activity levels. Figure 13 shows a block diagram of the convolutional layer.

## 7.2 Pooling Layer (PL)

The pooling layer is another type of feature extraction neural networks. The pooling layer decreases the size of the obtained HR image. It combines neighboring pixels of a certain area of the HR image into a single representative value. The used value is the maximum or mean value of the pixels. The max-pooling layer is used in the proposed technique. Figure 14 shows an example of the maximum pooling layer.

## 7.3 Rectified Linear Unit (ReLU)

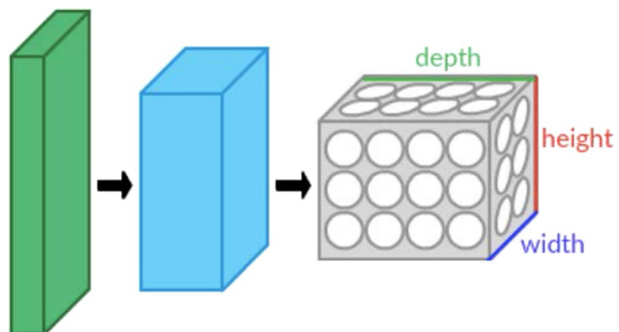
It allows faster and more effective training by mapping negative values to zero and maintaining positive values. The ReLU layer applies a linear activation function to the neuron output as shown in Eq. (67).

$$f(x) = \max(0, x). \quad (67)$$

## 7.4 Fully-Connected (FC) Layer

This layer reduces the size of input data to the size of classes that the CNN is trained for by combining the outputs of the CNV layer with different weights. Each neuron at the output of the CNV layer will be connected to all other neurons after being weighted properly. Similar to the CNV layer, weights of the taps in the FC layer are found through the backpropagation algorithm.

Fig. 13 Convolutional layer





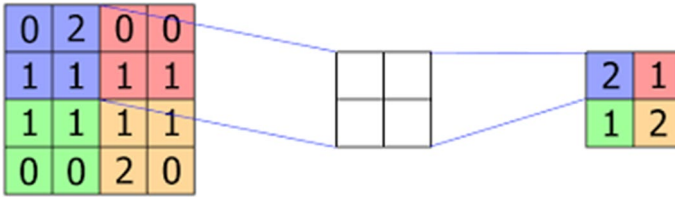


Fig. 14 Maximum pooling layer

### 7.5 Classification Layer

This is the final layer of the CNN that converts the output of the FC layer to a probability of each object being in a certain class. Typically, soft-max type of algorithms is used in this layer to provide the classification output as shown in Eq. (68).

$$P(y = j|\mathbf{x}) = \frac{e^{\mathbf{x}^t w_j}}{\sum_{k=1}^K e^{\mathbf{x}^t w_k}} \tag{68}$$

The proposed model consists of 3 CNV layers followed by 3 max pooling layers. Finally, a global average pooling is used. Figure 16 shows an exemplary architecture of the convolutional neural network. Images are input with size  $224 \times 224$ . Layers have numbers of filters of 16, 32 and 64 for layers 1, 2 and 3, respectively. Finally, a dense layer with size of 2 is used for the classification decision as shown in Figs. 15 and 16.

### 8 Performance Metrics

In this proposed approach, accuracy is used to estimate the strength of the CNN model. Accuracy is calculated as follows:

$$\text{Accuracy} = \frac{\text{No. of predicted images with tumors}}{\text{Total No. of HR images}} \times 100. \tag{69}$$

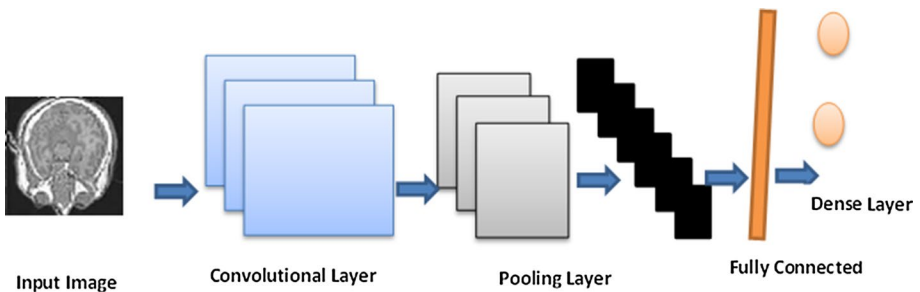


Fig. 15 Block diagram of the proposed classification approach

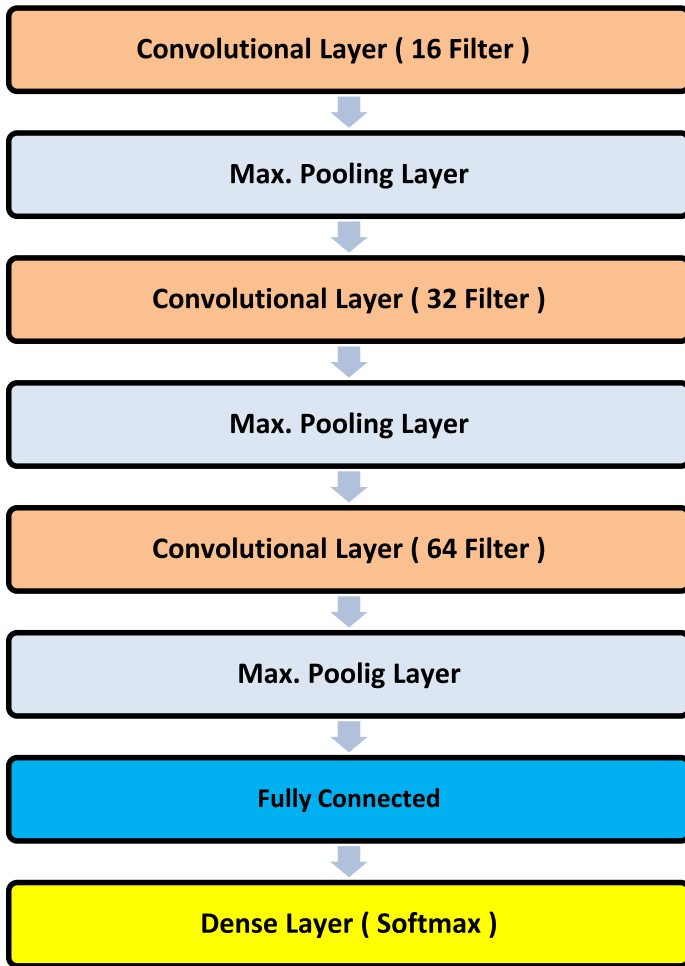


Fig. 16 Layers of the proposed deep learning classification model

## 9 Simulation Results

Simulation experiments are performed on a dataset of MR and CT images using python 3.5 [24], Tensorflow [25] and Keras [26]. This simulation is carried out on about 370 images. These images are the results of image fusion and interpolation techniques. Figure 17 shows the CT and MR images on which the image fusion and interpolation techniques are performed. Figures 18, 19, 20 and 21 show the images after fusion and interpolation processes for both normal and tumor states.

Figures 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32 and 33 show observations of accuracy and loss along the training epochs. Ten epochs and a batch size of ten have been set to train the deep learning model. Table 1 shows the deep learning model accuracy of all scenarios. These results show that image fusion based on wavelet transform combined with spline interpolation has the optimum accuracy of 90.9%. Table 1 gives a summary of accuracy levels in all scenarios.

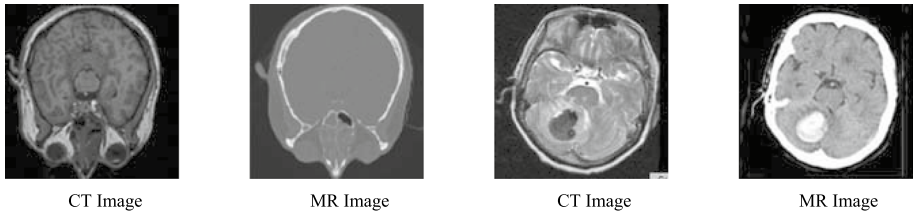


Fig. 17 CT and MR images of normal and tumor states

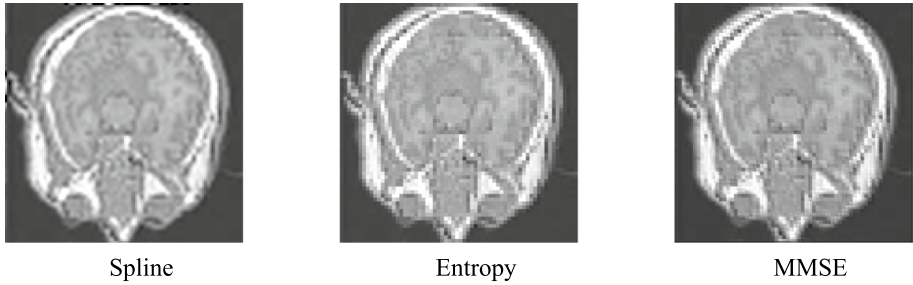


Fig. 18 Images after applying wavelet transform for image fusion with interpolation techniques (normal images)

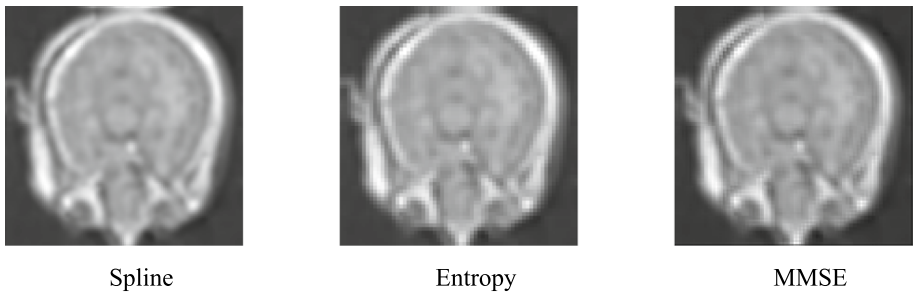


Fig. 19 Images after applying curvelet transform for image fusion with interpolation techniques (normal images)

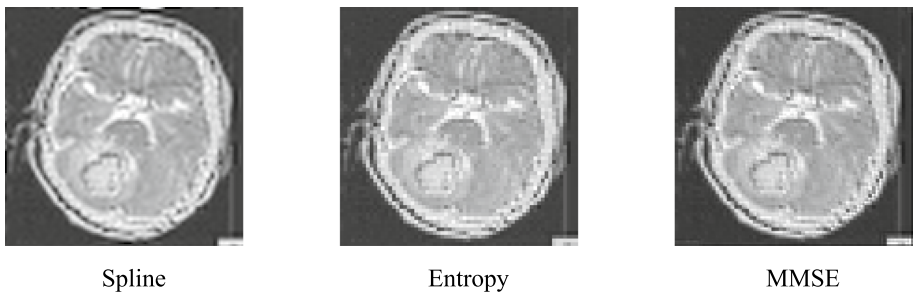
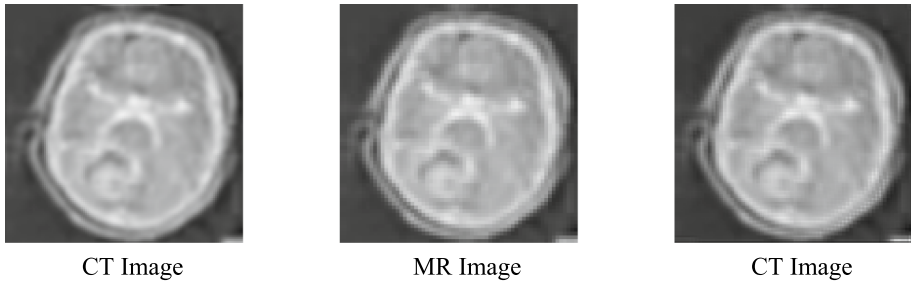
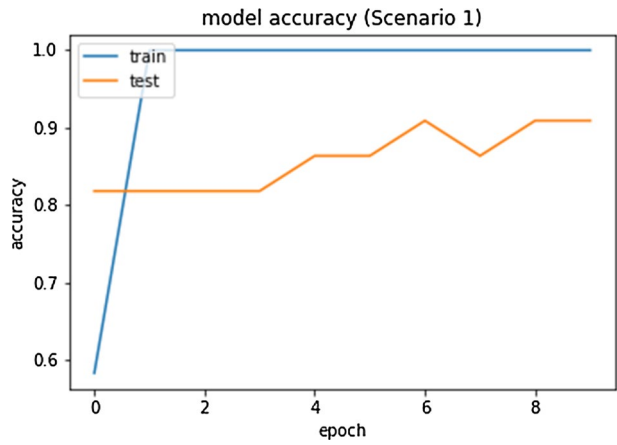


Fig. 20 Images after applying wavelet transform for image fusion with interpolation techniques (images with tumor)



**Fig. 21** Images after applying curvelet transform for image fusion with interpolation techniques (images with tumor)

**Fig. 22** Accuracy of Scenario #1



**Fig. 23** Loss of Scenario #1

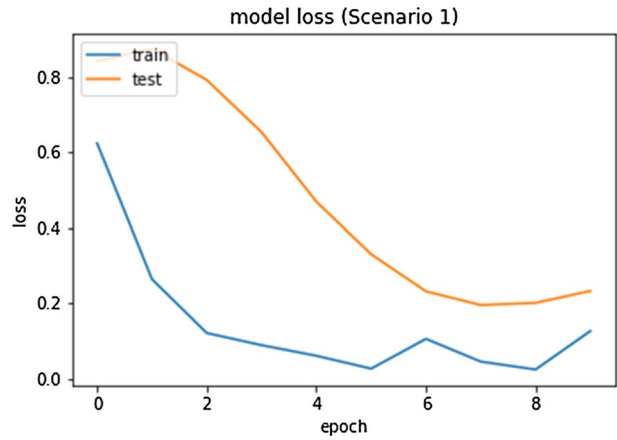


Fig. 24 Accuracy of Scenario #2

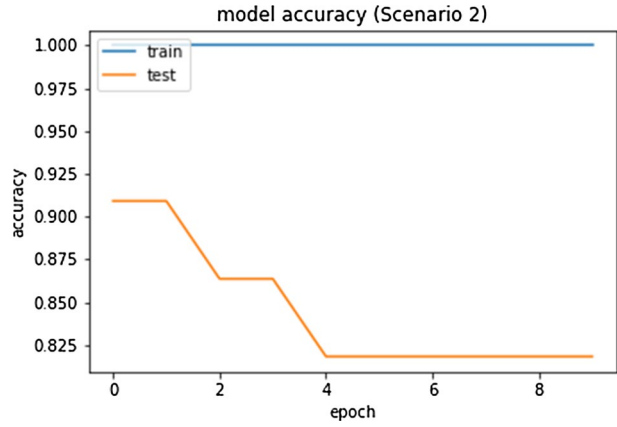


Fig. 25 Loss of Scenario #2

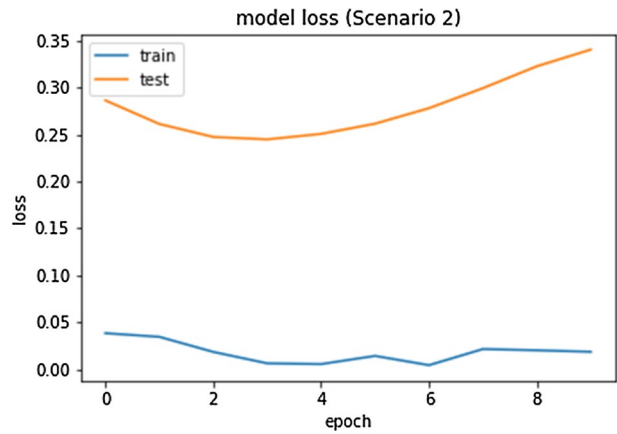


Fig. 26 Accuracy of Scenario #3

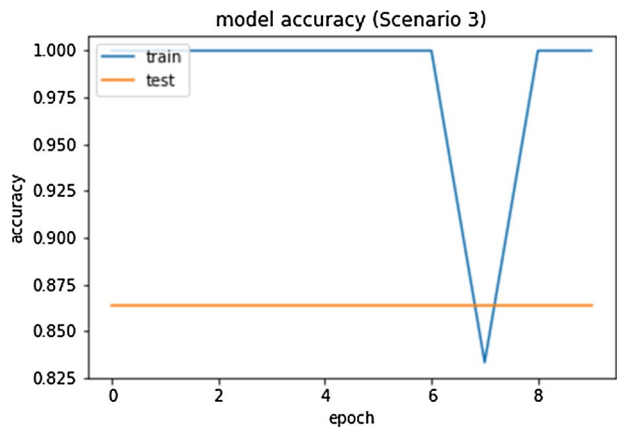


Fig. 27 Loss of Scenario #3

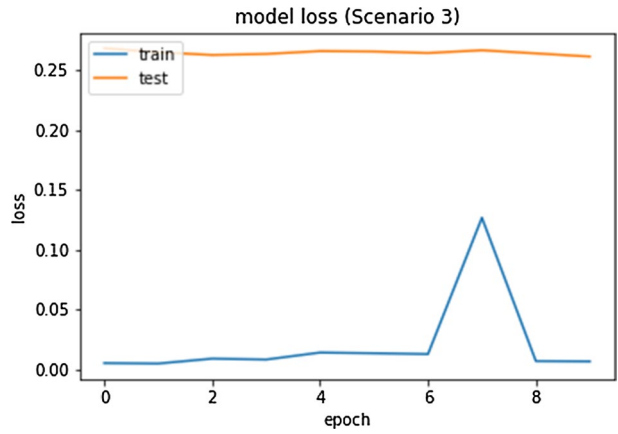


Fig. 28 Accuracy of Scenario #4

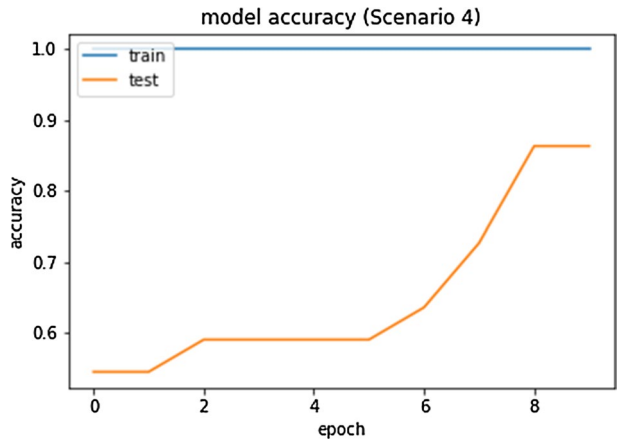


Fig. 29 Loss of Scenario #4

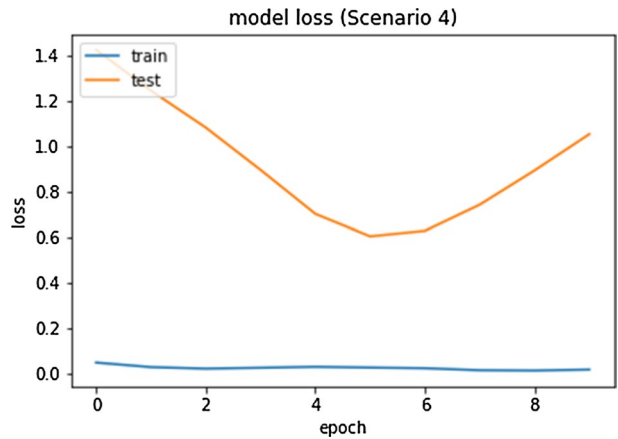


Fig. 30 Accuracy of Scenario #5

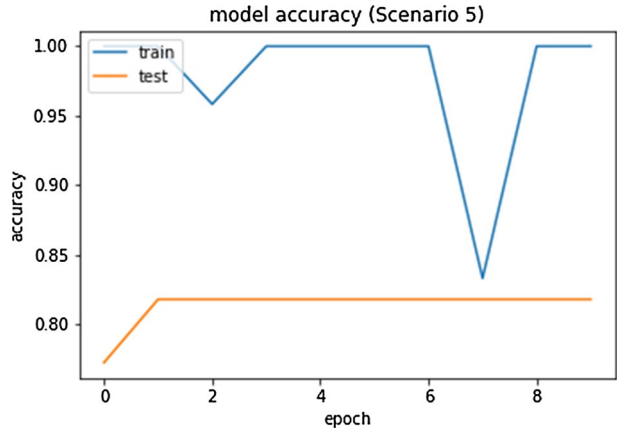


Fig. 31 Loss of Scenario #5

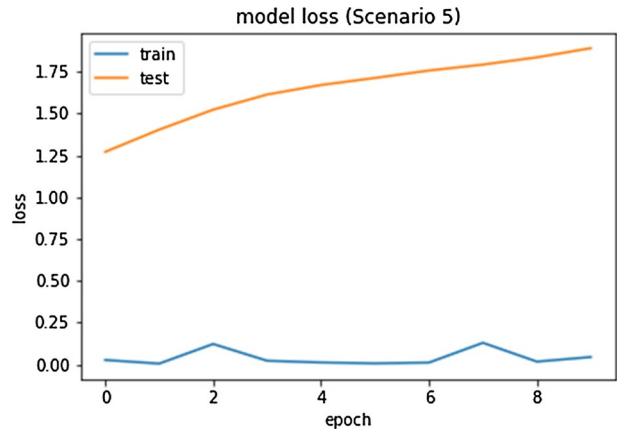
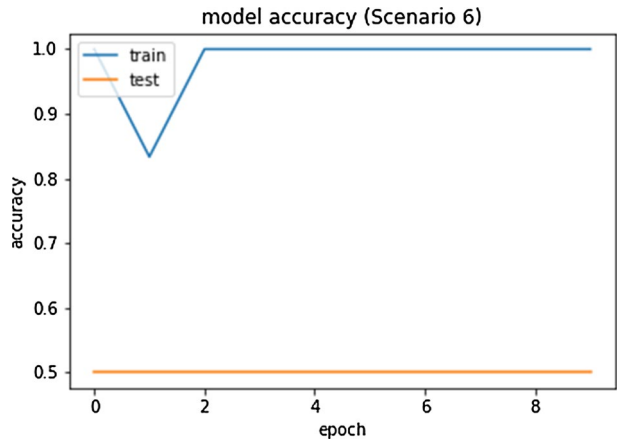
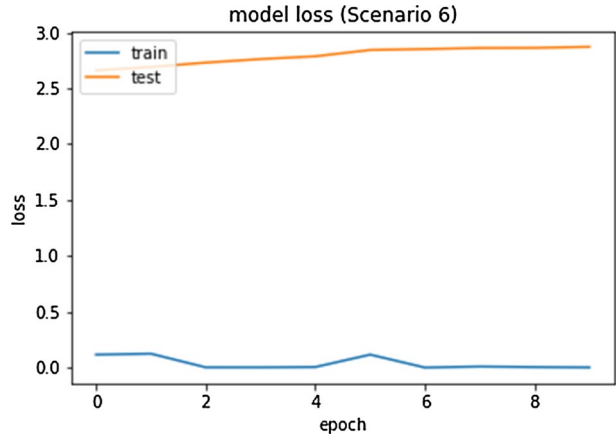


Fig. 32 Accuracy of Scenario #6



**Fig. 33** Loss of Scenario #6



**Table 1** Summary of accuracy levels of all scenarios

Image fusion	Interpolation techniques		
	Spline (%)	Entropy (%)	MMSE (%)
Wavelet transform	90.9091	81.8182	83.3636
Curvelet transform	86	81	50

The scenarios considered are:

- Scenario #1: Wavelet fusion with cubic spline interpolation.
- Scenario #2: Wavelet fusion with maximum entropy interpolation.
- Scenario #3: Wavelet fusion with LMMSE interpolation.
- Scenario #4: Curvelet fusion with cubic spline interpolation.
- Scenario #5: Curvelet fusion with maximum entropy interpolation.
- Scenario #6: Curvelet fusion with LMMSE interpolation.

## 10 Conclusion

Automated medical image diagnosis is the main objective of medical imaging that has not yet been achieved. This objective is considered in this paper with a framework comprising multi-modality image fusion, image interpolation, and deep learning classification. The issue of multi-modality image fusion has been considered with two different fusion algorithms: wavelet fusion and curvelet fusion. Different interpolation schemes have also presented and studied. Finally, the new trend of deep learning have been implemented for tumor detection. Accuracy levels of detection and automated classification up to 90% have been obtained. These results are promising in the field of automated medical diagnosis and can open the door towards more accurate classification results and efficient utilization of automated diagnosis in other medical applications.

**Acknowledgements** This research was funded by the Deanship of Scientific Research at Princess Nourah bint Abdulrahman University through Fast-track Research Funding Program.



## References

1. Diagnostic radiology physics: A handbook for teachers and students, ISBN:978-92-0-131010-1 © IAEA (2014).
2. Suetens, Paul. (2009). *Fundamentals of medical imaging* (2nd ed.). Cambridge: Cambridge University Press.
3. Ali, F. E., El-Dokany, I. M., Saad, A. A., & Abd El-Samie, F. E. (2008). Curvelet fusion of MR and CT images. *Progress in Electromagnetics Research C*, 3, 215–224.
4. Ali, F. E., El-Dokany, I. M., Saad, A. A., & Abd El-Samie, F. E. (2010). A curvelet transform approach for the fusion of MR and CT images. *Journal of Modern Optics*, 57(4), 273–286.
5. Ali, F. E., El-Dokany, I. M., Saad, A. A., Al-Nuaimy, W., & Abd El-Samie, F. E. (2010). High resolution image acquisition from MR and CT scans using the curvelet fusion algorithm with inverse interpolation techniques. *Journal of Applied Optics*, 49(1), 114–125.
6. El-Hoseny, H. M., Abd El-Rahman, W., El-Rabaie, E. M. (2018). An efficient DT-CWT medical image fusion system based on modified central force optimization and histogram matching. *Infrared Physics and Technology*, 94, 223–231.
7. El-Hoseny, H. M., Abd El-Rahman, W., El-Shafai, W., El-Rabaie, E. M., Mahmoud, K. R., & Abd El-Samie, F. E. (2018). Optimal multi-scale geometric fusion based on non-subsampled contourlet transform and modified central force optimization. *International Journal of Imaging Systems and Technology*, 29, 4–18.
8. Mahesh, & Subramanyam. M. V. (2012). Automatic feature based image registration using SIFT algorithm. In *IEEE third international conference on computing, communication and networking technologies (ICCCNT'12)*.
9. Zhu, Y., Cheng, S., Stanković, V., & Stanković, L. (2013). Image registration using BP-SIFT. *Journal of Visual Communication and Image Representation*, 24(4), 448–457.
10. Jie, Z. (2016). A novel image registration algorithm using SIFT feature descriptors. In *IEEE international conference on smart city and systems engineering (ICSCSE)*.
11. Soliman, R. F., El-Banby, G. M., Algarni, A. D., El-Sheikh, M., Soliman, N. F., Amin, M., & Abd El-Samie F. E. (2018). Double random phase encoding for cancelable face and iris recognition. *Applied Optics*, 57(36), 10305–10316.
12. Ashiba, H. I., Mansour, H. M., Ahmed, H. M., El-Kordy, M. F., Dessouky, M. I., Fathi, E., et al. (2018). Enhancement of infrared images based on efficient histogram processing. *Wireless Personal Communications*, 99(2), 619–636.
13. Ashiba, H. I., Mansour, H. M., Ahmed, H. M., Dessouky, M. I., El-Kordy, M. F., Zahran, O., & Abd El-Samie, F. E. (2018). Enhancement of IR images using histogram processing and the undecimated additive wavelet transform. In *Multimedia tools and applications* (pp. 1–14).
14. El-Khamy, S. E., Hadhoud, M. M., Dessouky, M. I., Sallam, B. M., & Abd El-Samie, F. E. (2005). Efficient implementation of image interpolation as an inverse problem. *Journal of Digital Signal Processing*, 15(2), 137–152.
15. El-Khamy, S. E., Hadhoud, M. M., Dessouky, M. I., Salam, B. M., & Abd El-Samie, F. E. (2006). Efficient solutions for image interpolation treated as an inverse problem. *Journal of Information Science and Engineering*, 22, 1569–1583.
16. El-Khamy, S. E., Hadhoud, M. M., Dessouky, M. I., Salam, B. M., & Abd El-Samie, F. E. (2008). New techniques to conquer the image resolution enhancement problem. *Progress in Electromagnetics Research B*, 7, 13–51.
17. Thevenaz, P., Blu, T., & Unser, M. (2000). Interpolation revisited. *IEEE Transactions on Medical Imaging*, 19(7), 739–758.
18. Blu, T., Thevenaz, P., & Unser, M. (2001). MOMS: Maximal-order interpolation of minimal support. *IEEE Transactions on Image Processing*, 10(7), 1069–1080.
19. Unser, M. (1999). Splines A perfect fit for signal and image processing. *IEEE Signal Processing Magazine*, 16(6), 22–39.
20. Nishanth, K., & Karthik, G. (2015) Identification of diabetic maculopathy stages using fundus images. *Journal of Molecular Imaging & Dynamics*. <https://doi.org/10.4172/2155-9937.1000118>
21. LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2324.
22. Ranzato, M. A., Huang, F. J., Boureau, Y. L., & LeCun, Y. (2007). Unsupervised learning of invariant feature hierarchies with applications to object recognition. In *IEEE conference on computer vision and pattern recognition, 2007. CVPR'07* (pp. 1–8). IEEE.

23. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1), 1929–1958.
24. <https://www.python.org/downloads/release/python-350/>. Accessed 11 July 2018.
25. <https://www.tensorflow.org/>: An open source machine learning framework for everyone. Accessed 11 July 2018.
26. <https://keras.io/>: The Python Deep Learning library. Accessed 11 July 2018.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Abeer D. Algarni** received the B.Sc.(Hons.) in Computer Science from King Saud University, Riyadh, Saudi Arabia in 2007. She received the M.Sc., and Ph.D. degrees from School of Engineering and Computer Sciences, Durham University, United Kingdom in 2010 and 2015, respectively. She worked as an assistant professor at the College of Computer and Information Sciences at Princess Nourah Bent Abdulrahman University since 2008 up till now. Her current research interests include networking and communication systems, digital image processing, digital communications and cyber security.