Classification of Medical Images using Deep and Handcrafted Visual Feature-based Algorithm

M.Jaiganesh [1], K.V. Archana [2], J.Jeganathan [3], G.N. Balaji [4], Dr.R.Nagarajan [5], Dr.P.Jenopaul [6]

[1, 2, 4] Faculty of Engineering and Technology, Jain University, Bangalore,
Karnataka, India.
Email Id: jaidevlingam@gmail.com, archanaglows@gmail.com,
balaji.gnb@gmail.com

[3] PSNA College of Engineering and Technology, Dindigul, Tamilnadu,
India.
Email Id: Jeganathan7000@gmail.com

[5] Professor, Department of Electrical and Electronics Engineering,
Gnanamani College of Technology,Namakkal-637018. Email Id: krnaga71@yahoo.com

[6] Professor, Department of Electrical and Electronics Engineering,
Adi Shankara Institute of Engineering and Technology, Kerela-683574.
Email Id: jeno.eee@adishankara.ac.in

Abstract: Though the medical images contain a lot of information, there is a problem in retrieving that information and make it useful for further diagnosis. Information from these medical images can be utilized effectively by implementing classification and retrieval. Features of the image are the most important factors for the image classification. Different kinds of handcrafted features are available which include single descriptors for colour, texture, and shape and combined descriptors. Different feature extraction algorithms like LBP and BOF are used to extract these features. We also have many deep learning techniques that extract deep learned features and are widely acknowledged as a powerful tool for image classification. But due to lack of not large enough dataset, over-fitting may occur. To address these problems, a combined deep and handcrafted visual feature-based algorithm is implemented in this work.

1 Introduction

In recent times, there is a huge increase in digitalization of medical imaging which in turn increased the number of medical images rapidly. Medical imaging is a useful resource and is an invaluable tool for clinicians since it is a process where the interior of a body and its function is visually represented for analysis and intervention that seeks to reveal internal structures. As the medical images act as a key source of information in medical processes such as diagnosis of diseases and surgical planning, the research area of medical image classification is very active from past decade many methods are being introduced. Usually, the labelling of these images is performed manually and needs a lot of professional expertise, which is time-consuming, and prone to errors because of human subjectivity and variable image quality. This led to the dire need for automatic medical image classification. Different sorts of medical imaging techniques such as X-ray, MRI (Magnetic Resonance Imaging), CT (Computer Tomography), Ultrasound, etc. help constrict the

causes of an injury and make sure that the diagnosis is accurate. X-rays are most widely used amongst these techniques. Although you need more subtle tests, you are likely to get an x-ray first.

The fundamental step for X-ray image classification is to extract meaningful features. Depending upon the feature selection the accuracy of classification varies. So, the choice of features plays a key role. Especially in the medical domain, there should be no chance of error. So, the accuracy of classification should be at its best. There are various feature extraction techniques available for extracting significant features from an image. Existing image features can be categorized as Handcrafted features and Learned features. The proposed system involves classification of X-ray images using both handcrafted visual features and deep learned features extracted from pre-trained CNN.

2 Related Work

A classification is a form of the data analysis process that extracts models describing important data classes. The classification has numerous applications that include credit approval, performance prediction, manufacturing, fraud detection, target marketing, and medical diagnosis. Classification is a two-step process, consisting of a learning step and a classification step. Classification can be performed on any kind of data such as Text, Numerical, Audio, Video, Images.

By using classification, images can be mapped to one of several predefined classes. This process is also called as Automatic Image Categorization. In many examples where image classification has been successfully implied it can be argued that the classification was fairly straight forward in the sense that the task could easily have been conducted by humans. For some applications, the support provided by image classification techniques is more essential in that the image data considered cannot be readily categorised by human interpretation. One area of application where this is the case can be found in the domain of medical imaging, where differences between images associated with different class labels are sometimes hardly noticeable. Image classification in this latter case represents a more challenging task. Large Image data repository and image quality pose major challenges for image classification, review and assimilation for clinical care and research.

Feature extraction involves the extraction of significant features from a huge detail. Several feature extraction techniques are available for extracting significant features from an image. Existing image features can be broadly divided into two categories, namely, Hand Crafted features and Learned features.

Handcrafted features [1] are manually extracted from images according to a certain predefined algorithm based on expert knowledge. Depending on the problem different handcrafted features are to be used. Local Binary Patterns (LBP), Bag of Features (BOF), Average Grey Descriptor (AGD), and Grey Level Co-occurrence Matrix (GLCM) are some of the examples of handcrafted features [2].

In Jeanne's work [2], Firstly feature extraction was performed where the performance of five different feature types including Average Gray Descriptor (AGD), Colour Layout Descriptor (CLD), Edge Histogram Descriptor (EHD), Gray-level Co-occurrence Matrix (GLCM) and Local Binary Patterns (LBP) was investigated. Local binary patterns achieved better performance values and outperformed other feature types.

Local Binary Patterns (LBP) [2] is a greyscale local texture descriptor. The LBP operator labels image pixels by thresholding a neighbourhood of each pixel with the centre value. Initially, for each pixel, its eight neighbours are first followed along a circle and the pixel value is used to threshold each neighbours value, resulting in eight binary numbers. Then an 8-bit encoded is formed by

concatenating these binary numbers. Finally, the histogram of the frequency of each integer occurring over the entire image is counted as 256-dimensional LBP descriptor.

Bag of Features (BOF) [3] is considered as a vector of occurrence counts of local descriptors over a visual vocabulary. This process consists of three steps, initially, the Scale Invariant Feature Transform (SIFT) algorithm is used for detecting key points and generate 128-dimensional descriptor for each key point. Then vector quantization is used to assign SIFT descriptors to clusters. Then finally the distribution of SIFT descriptors over the visual vocabulary is counted as the BOF descriptor.

Histograms of Oriented Gradients (HOG)[4] is a global descriptor. HOG features are used to discriminate body shapes. Initially, image is divided into patches then their gradient orientation, weighted by their gradient magnitude is calculated and considered as HOG values. Then these values are represented as a single vector to serve as a HOG feature. One advantage of HOG is its robustness against illumination variance.

Mueen et al. [5] proposed the classification of X-ray images performed by combining global, local and pixel features in one big feature vector. In the first stage, Global features are extracted, where it failed to get a good accuracy rate for some classes although those classes have enough training images. In the second stage, Local level features are extracted and can distinguish better than global features for similar classes. In the third stage, Pixel level information was extracted and provided results for classes with fewer training images, which was a failure in the case of global and local features. In the final stage, all these three levels are combined into one feature vector which has a high dimension. To avoid memory and runtime problems, it was reduced by using PCA.

Zare et al. [6] proposed an approach where the automatic classification of medical X-ray images was performed using different feature extraction techniques such as Gray Level Co-occurrence Matrix (GLCM), Canny Edge Operator, Local Binary Pattern (LBP), pixel value as low-level image representation, and Bag of Words (BoW) as local patch-based image representation. These features have been exploited in different algorithms for classification. Performance obtained was analysed regarding the image representation techniques used and results showed that LBP and BoW outperformed the other algorithms.

3 Proposed Methodology

In this proposed work, for classification of X-ray images two-level classification was performed.
• In the first level, three different features were used for predicting the correct output.
• By using those three predictions, the final prediction was done.

The combined approach of Deep and Handcrafted classification for extracting features of x-ray images classification is shown in figure 1.
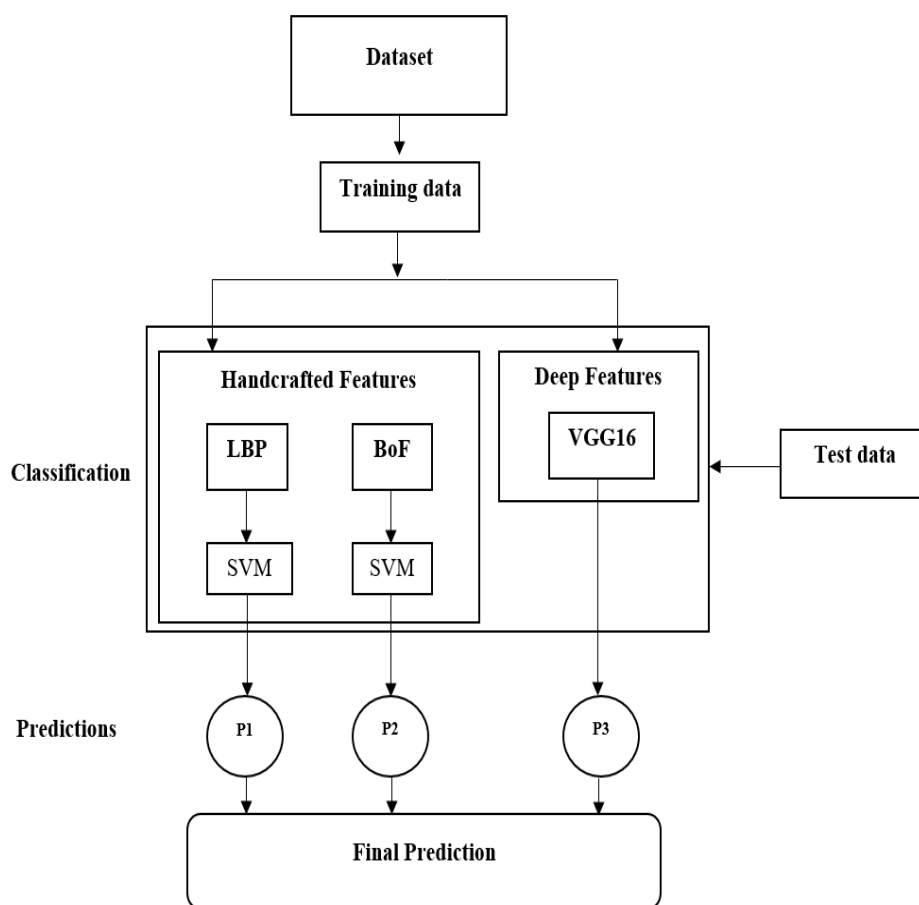
Fig. 1 System Architecture

3.1. Pseudo Code

Step 1: The downloaded MURA dataset was first separated into training and testing sets.

Step 2: Initially, the following three models were trained on all the images of the training set separately:

1. LBP + SVM Classifier
2. BOF + SVM Classifier
3. VGG 16

Step 3: Here, Local Binary Patterns (LBP) and Bag of Features (BOF) comes under the category of handcrafted   feature extraction models, whereas VGG16 is used as a deep Convolutional neural network model to extract deep features.

Step 4: In LBP,
- Each X-ray image in the training set is first divided into blocks of the same size.
- Then calculate the LBP code for each pixel in a block by comparing the central pixel with its 8 neighbours.
- If the neighbour's value is greater than the value of a central pixel, then write '1', and if the neighbour's value is less than the value of a central pixel, then write '0'. Follow this for all pixels of a block in a clockwise direction.
- This results in an 8-digit binary number which is then converted into the decimal, yielding a

value i.e. LBP code.

- For every block, compute the histogram over the output LBP array.
- These histograms are then concatenated into a single feature histogram as the description of the x-ray image.
- The feature vector can now be processed using the Support Vector Machine (SVM) classifier where the training of the model is done.

Step 5: In BOF,

- Initially, a dictionary of visual words is created which is called a bag of features. The local interest points are to be identified in an X-ray image. Scale-Invariant Feature Transform (SIFT) extracts key points and computes their descriptors. Scale-invariant feature transform (SIFT) is one of the famous feature descriptors that can handle the rotations, scale, intensity, and affine variations.
- The features extracted are to be clustered using K-means clustering. Each feature cluster of similar image descriptors extracted represents a visual word.
- The final feature vector is a histogram counting the occurrences of each visual word in the image.
- Finally, after the histograms are generated, they were sent to the SVM Classifier along with the labels for the training.

Step 6: In VGG16,

- Firstly, the pre-trained VGG16 model was loaded from Keras applications without the output layer.
- Freeze all the layers and then a new model was created to which this pre-trained model was added as a base model.
- Add new layers to the model.
- Two separate Image Data Generators are used to read all the images in the training and test datasets.
- Before training, compile the model for configuring the learning process.
- Then the learning rate is set and training was performed. After training, the model was saved.

Step 7: To do the predictions on the trained model, saved models of LBP, BOF and VGG need to be loaded.

Step 8: In the final classification phase, a new X-ray image is passed as the input, where three predictions from the three models resulted.

Step 9: At last, based on the above three predictions, a final prediction has resulted.

3.2 Dataset

MURA (**Mu**sculoskeletal **Ra**diographs) [7], provided by Stanford University School of Medicine is a large dataset of bone X-rays. MURA is one of the largest public X-ray image datasets with a total of 40,561 multi-view musculoskeletal radiographic images. Each image in the dataset belongs to one of the seven classes: elbow, finger, forearm, hand, humerus, shoulder, and wrist.

3.3 Handcrafted Features

Local Binary Patterns (LBP) and Bag of Features (BoF)/Bag of Visual Words (BoVW) belong to a non-automatic feature extraction method. These are used as handcrafted features for this project and are used to train models separately using Support Vector Machine classifier.

Local Binary Patterns

Local Binary Patterns (LBP) was used as one of the handcrafted features in this project because the medical images are the greyscale images. Local Binary Pattern introduced by Ojala et al [8] is an efficient texture operator that labels the image pixels by thresholding the 3 x 3 neighbourhood of each pixel with the centre value and considering the result as a binary number (also known as LBP Code).

Bag of Features

The Bag of Features (BoF), originated from Bag of Visual Words (BoVW), is characterized as an order-less collection of image features. This method learns meaningful features and describes images in terms of the histogram of these features.

3.4 Deep Features

One important aspect of Convolutional Neural Networks (CNN) is that they can automatically learn hierarchical feature representations. This means that features computed by the early layers are general and can be reused in different problem domains, while features in the last layer are more dataset-specific and depend on the chosen task. Using a pre-trained Convolutional neural network for classification is a lot better than to build a CNN from the scratch.

When we train a network from scratch, we encounter the following limitations:

• Since the network has millions of parameters, to get an optimal set of parameters, we need to have huge data.
• Even if we have a lot of data, training generally requires multiple iterations and it takes a toll on the computing resources. Hence, a lot of computing power is required

Thus, fine-tuning a network avoids these limitations and helps to tweak the parameters of an already-trained network so that it adapts to the new task at hand. The initial layers learn very general features and the later layers tend to learn patterns more specific to the task it is being trained on.

VGG16 (Visual Geometry Group)

In this paper, pre-trained VGG16 model was used where the transfer learning was performed to it. VGG16 is one of the successful Convolutional neural networks where it has a deeper architecture. It was trained on more than a million images from the ImageNet database [9]. The Keras Applications module has pre-trained deep learning models that have pre-trained weights trained on ImageNet. The reason why transfer learning works so well is that we use a network which is pre-trained on the ImageNet dataset and has already learnt to recognize the trivial shapes and small parts of different objects in its initial layers. By using a pre-trained CNN to do transfer learning, we are simply adding a few dense layers at the end of the pre-trained VGG model and learning what combination of these already learnt features help in recognizing the objects in our new dataset. Therefore, we train only a few dense layers and additionally, we are using already learnt trivial

features to recognize new objects. This helps in making the training process very fast and requires very less training data compared to training a Convolutional network from scratch. In Fine-tuning, freeze the already trained low-level features, and only train high level features needed for our new image classification problem. Fine-tuning was performed according to the X-ray images dataset.
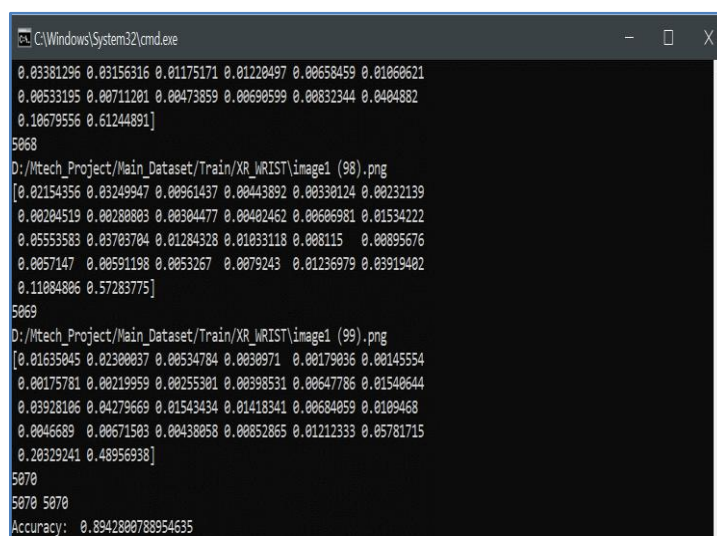
4   Experimental Results and Discussion

In this phase three models LBP+SVM, BOF+SVM and VGG16 were trained separately with the above-segregated training dataset [10]. Once the MURA dataset consisting of X-ray images of 7 human body parts is downloaded, it is arranged into 7 different classes where each class consists of each body part. Then the dataset (MURA) is split into two folders i.e. training and validation sets. While training dataset is used to build and train the model, the validation dataset is used to validate and test the model accuracy [14]. Both the training and validation sets have seven individual folders of images representing seven different categories of body parts labelled as XR_ELBOW, XR_FINGER, XF_FOREARM, XR_HAND, XR_HUMERUS, XR_SHOULDER, and XR_WRIST.

4.1 Local Binary Pattern + SVM Classifier

LBP (Local Binary Patterns) is a gray scale local texture descriptor. Each X-ray image in the training dataset is divided into blocks. [11] For each block, LBP histogram was generated using the neighbourhood of every pixel with the centre value and finally, all the histograms were appended to form a single feature histogram.

Feature histogram of every X-ray image with its label is sent to the SVM classifier where the training of the model is done. It can be observed that the model trained using Local Binary Patterns has achieved 89% of accuracy for the training set, as shown in figure 2.

Fig.2 Training accuracy of LBP model

Fig. 3 Testing accuracy of LBP model

From figure 3, it can be observed that the model trained using the Local Binary Patterns has achieved 84% of accuracy for the testing set.

4.2 Bag of Features + SVM Classifier

Initially, SIFT features were extracted from every image of the training dataset. Then the Bag of Features was built by reducing the number of features using k-means clustering. It can be observed that the model trained using Bag of Features has achieved 98% of accuracy for the training set. It can be observed that the model trained using the Bag of Features algorithm has achieved 95% of accuracy for the testing set.

4.3 VGG16

Firstly, the pre-trained VGG16 model was loaded from Keras applications.
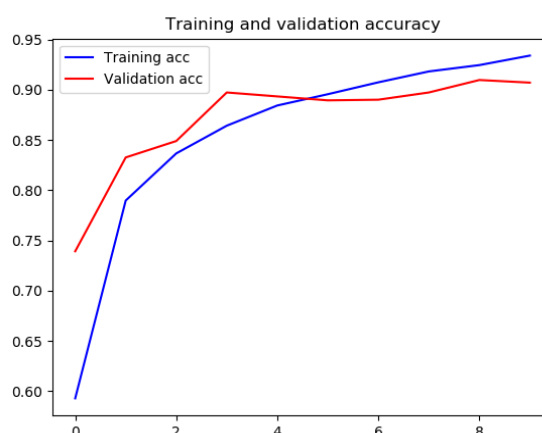
Fig.4 Training and Validation accuracy plot of VGG16

It can be observed that, from figure 4, by fine-tuning the pre-trained VGG-16 using transfer learning, the model has achieved 93% of training accuracy and 90% of validation accuracy.

5 Conclusion

A body part was classified from the X-ray images. Both handcrafted features and learned features were used in a combination approach. LBP and BOF were used as handcrafted features and were trained separately using an SVM classifier. VGG16 was considered as pre-trained CNN and fine-tuning was performed. The pre-trained DCNNs were able to transfer the knowledge of image representation learned from natural image datasets to specific image classification tasks. The two-level classification was implemented, where LBP+SVM, BOF+SVM and VGG16 were used at the first level. In Fusion method, majority voting classifier predicted the accurate result of the body part.

References

[1] Antipov, G., Berrani, S., Ruchaud, N., Dugelay, J.: Learned vs. Hand-Crafted Features for Pedestrian Gender Recognition. In: Proceedings of the 23rd ACM international conference on Multimedia - MM '15. pp. 1263–1266 (2015)

[2] Jeanne, V., Unay, D., Jacquet, V.: Automatic detection of body parts in x-ray images. In: 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. 25-30 (2009)

[3] J. Zhang, Y. Xia, Y. Xie, M. Fulham and D. D. Feng, "Classification of Medical Images in the Biomedical Literature by Jointly Using Deep and Handcrafted Visual Features," in IEEE Journal of Biomedical and Health Informatics, vol. 22, no. 5, pp. 1521-1530, Sept. 2018.

[4] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, USA, 2005, pp. 886-893 vol.1.

[5] Mueen, A., Sapiyan Baba, M., Zainuddin, R.: Multilevel Feature Extraction and X-Ray Image Classification. Journal of Applied Sciences 7(8), 1224–1229 (2007)

[6] Zare, M. R., Seng, W. C., Mueen, A.: Automatic Classification of medical X-ray Images. Malaysian Journal of Computer Science. 26(1). 9-22 (2013)

[7] https://stanfordmlgroup.github.io/competitions/mura/

[8] Ojala, T., Pietikäinen, M., Harwood, D.: A comparative study of texture measures with classification based on featured distributions. Pattern Recognition. 29, 51-59 (1996)

[9] ImageNet. http://www.image-net.org

[10] Tejaswini Reddy Naini, M. Jaiganesh, S. Suguna Mallika, Suchith Buddha.: Classification of X-ray Images for Human Body Parts, International Journal of Psychosocial Rehabilitation, 24(8), 12839-12844 (2020).

[11] Dr.M.Meenakumari,Dr.T.Mohanasundaram,R.SureshKumar,A.MariaSindhuja,S.Gowdhamkumar,An Efficient Method for Text Detection and Recognition in Still Images, Annals of the Romanian society for cell biology, ISSN: 1583-6258, Vol. 25, Issue 3, 2021, Pages. 7408-7415.