



A Survey on Tagged Image File Format (TIFF)

U. V. Ram Mohan Rao ^{a}, R. Sai Manikanta ^a, V. Durga Prasad ^a, S. Modha Madhan ^a, Y. V. S. Kalyan ^a*

Department of Computer Science and Engineering, GMR Institute of Technology, Rajam, Andhra Pradesh, India.

ABSTRACT

This paper presents the literature on Tagged file format (TIFF), which is called as Tag Image File Format. It is a computer record used to store raster designs and picture data. TIFF are a helpful method for putting away great pictures prior to altering if you have any desire to keep away from loss document designs. In this paper the survey was taken on the medical images, image segmentation, reorganization and compression. Tiff format is used for handling the high quality of the image file. This survey is carried on the various method and technologies like machine learning algorithms, deep learning techniques and image segmentation. The main focus of this survey is on how the various pre-processing techniques and the feature extraction techniques are used on the tiff images. In any image processing or medical image diagnosis the pre-processing and the feature extraction are the major part of the work so the we have taken this survey on the various architectures how they processed this high quality image without any loss to the useful or mandatory data present in the images.

Keywords: Convolution neural network, ResNet, Image Segmentation, Tailing

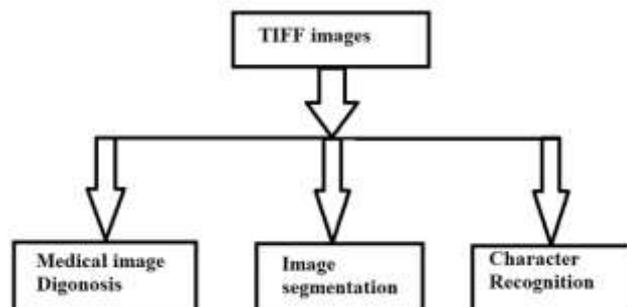
1. Introduction

Nowadays, Images are being the most important aspect in the disease diagnosis, image processing and segmentation. In medical image processing the images are being the most important while compression of the images with large size it's being the tough to preprocess the images so that when we are applying the lossless compression algorithms. When the images are in the large size it's difficult to train the model in the machine learning or deep learning and they occupy the large amount of memory to store so to overcome this type of problems we mainly use the image rescale or compression techniques but sometimes when we compress an image they may be a chance of losing the data from the image due to change in the size but to reduce the lose during the compression and rescaling we use the tagged image file format. This file format helps us to compression faster and the decomposing such file with lossless quality.

2. Work:

This study is proposed for the various methodologies and preprocessing techniques on the tagged file format images. The tagged file formatted images are main used in the medical imaging application, character reorganization and image segmentation. The tiff images are more advantageous for because they are capable for storing the high image quality which is more required for the medical images because the every single pixel in the medical image is most importing for the diagnosis of the various disease. These file format also used for the optical character recognition is able to transform scanned document into a editable and searchable text file. So that in this paper the survey is done on the various paper which they worked on the large size of images and tiff format images as the dataset and we have studied about how they are preprocessed the images and feature extraction techniques because the feature extraction and preprocessing are the main part in the any analysis.

3. TIFF images Classification:



The TIFF images are mainly classified for the medical image diagnosis, image segmentation and optical character recognition. Tiff images generate the wide range of band and each pixels. It indicates the pixel value of the image in the form of binary values which helps in the better classification. The medical images will contain the large data in the images so that tiff tagged images will be helpful in lossless compression. They are also used for the optical character reorganization techniques.

3.1 Medical images:

The dengue fever [3] is analysed based on the patient's plasma samples to identify the amount of virus that is infected to the patient. Spot blood vessels of finding and gridding were performed using Code Link Expression Analysis software. Each sample's digital TIFF tagged image file was transformed to a various number measuring hybridization intensity. Samples of each digital TIFF image file were transformed to a numerical value that measured hybridization intensity. Further analysis was on spot signals to identify performance ratio greater than one using the signal-to-noise. They were statistically analyzed, and the spot intensities are standardized claiming the median normalization approach and the Code Link microarray programmed. SAM (significance analysis of microarray) is a method for analyzing large-scale gene or protein expression data obtained with microarrays. The data is processed using this manner.

The cell loss recover scheme on tiff images to pre-processing and post processing techniques and comparison of the how the image quality is transmitted. "Because of their higher rate of transmission speed, Asynchronous Transfer Mode (ATM) networks are ideal for future integrated broadband service digital networks (BISDNs), one which incorporate the range of various services, that are including multimedia data delivery" [5]. TIFF files, which consist of the single image it will be spitted into equal half's distinct layouts are the most suitable format for the pre-processor since we can easily separate the image file transformed into essential and permissive parts by mentioning the front half mandatory. The initial sequence of the image pixels is reforming by over ride the results of method, which is same as the processor's single-cell loss case. The image has 75 pixels when it is compressed from a JPEG file to a TIFF file.

The diagnosis of active Tuberculosis (TB) prediction utilising Artificial Intelligence (AI) based automated detection of abnormalities in chest radiography is critical in radiology workflow. The data set, which includes 662 photos in tagged image file format (TIFF). The input data was pre-processed before being fed into the model, which used U-net architecture to segment the area of interest (ROI). It takes the most extreme points from the lung region ROI and paddings them with an extra 50 pixels to ensure the bounding box keeps the ROI. The Resnet50 and Dense net architectures were utilised in the article for model training, validation, and testing. The Resnet50 and Densenet121 input image sizes were set at 224x224. The classification was taken based on the Average pooling layer and Dense layer with ReLu activation. the accuracy metrics of the model are specificity as 0.98, sensitivity as 0.985. they have achieved remarkable calcification accuracy of 100%.

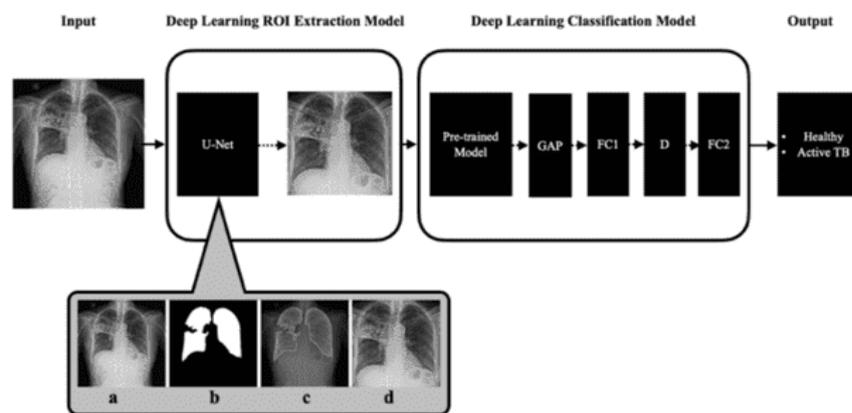


Fig .1 X-Ray of chest images using U-Net and pre-trained models Resnet50 and DenseNet121

This study diagnosis performed on the cell architecture of inflammation using histopathology images . The data set contains the 198 tiff images as the sample. In this they have used the DAPI channel images for tasks by UNet for segmentation and classification. This architecture used Gaussian Error Linear Unit (GELU) activation in all the layers. By applying the image tiling of 256 size, margin padding is initially empty, and some of tiles do not have any portion of the cell stained on them. The mean-squared error loss is identified by using the simple encoder-decoder architecture. This after processing auto encoder technique takes the output of the segmentation model as input from source and then computes this reconstruction of loss.

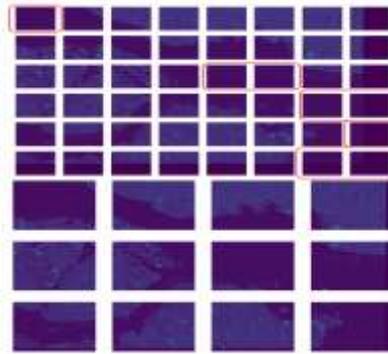


Fig.2 Image tiling reduces the number of blank tile (256*265 image size).

“TIFF files can be transformed into dual personality DICOM TIFF Digital Imaging and Communications in Medicine files without affecting the pixel data (DICOM). One or more images, each of which can be encoded as strips or tiles, can be encountered in TIFF files” [18]. To form tiled WS images that are in the form of TIFF and DICOM representations can share tiles with. No qualitative variations between both the origin TIFF image and the transformed DICOM-TIFF with dual-personality images were observed that's because the congested JPEG bits stream is not modified in process of conversion procedure. The strategy relies on the same pixel data organization in both formats the use of tiles and the use of the same compression supported in order to reduce the duplication of the whole image information. These applications considers the tiff tags while maintaining all of the benefits of the standard format can assist in overcoming the delay the progress toward the mainstream using WSI encoding of DICOM technique. Algorithm towards lossless medical imagery compression [19]. The various Transform techniques are built in such a way that the compression is maximized when there is a considerable amount of information. The study's purpose is to create a novel lossless compression model that exceeds the widely used TIFF and JPG graphic formats in medical image compression. Three datasets of medical imaging are presented. The initial dataset incorporates six hundred and twelve objects that correspond to original 24-bit colour colonoscopy pictures taken in tiff tag. The third dataset contains 517 original 24-bit colour knee X-ray pictures (1 and 2 knees). The third dataset consists of the five hundred and seventeen 24-bit colour knees of the X-ray images with two equal portions in original PNG format. Using CNN, they built the brand-new spectral with the transformation. The data reinforce the use of 2D TDC in medical imaging, of which its images consists huge information and therefore must take up as small space as willing for processing the data, display the information, and transmission. Overall accuracy of 2D TDC is 84.2%.

Colorectal cancer [9] is dignosised because it is having the high mortality rate as it is a dangerous disease. They selected a Unet-architecture network model since it has demonstrated outgrowth in segmentation of the medical image, incorporate the estimates the polyps from colonoscopy of the images. They proposed modifying the Residual Recurrent Unet as the architectural model to reduce model volume while maintaining methods performance. The dataset contains 612 tiff formatted images extracted from the 25 different colonoscopy images where most of them are in size as 384*288. Later the original image is rescaled to 224*224. In this the loss function handled with the include weighted cross-entropy (WCE) and the dice-loss. They have compared with the three datasets on this model CVC-ClinicDB with Dice Coefficient as 91.91, ETIS-LaribpolpDB with Dice coefficient as 81.62 and CVC-ColonDB with Dice coefficient as 90.09.

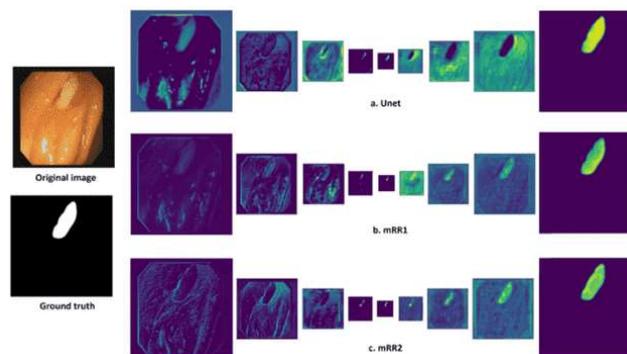


Fig.4 the demonstration of Unet, mRR1 and mRR2 techniques for the feature mapping.

Chest X-ray radiographic (CXR) [10] images are used for lung disease diagnosis. They employed the transfer learning strategy for the categories lung disorders using CXR pictures the increase model's effectiveness and computer-aided diagnostic systems (CAD) diagnostic as accuracy of the model performance. The SCH data set contains 35,282 images where the size of the are varying from 2000*2000 to 3000*3000 and the bit depth was 16 bit [0-65,535]. After the data augmentation the image size is resized to 1024*1024. They used CXR pictures (gray scale images) to classify multi-class lung illnesses without using data augmentation or transfer learning. They used EfficientNet v2 as a basic technique and defined the novel upper most layers as the fine-tuning techniques to perform the transfer learning technique. EfficientNet v2 model had obtained the 82.15% of accuracy.

Lossless compression techniques [17] have been embraced by the medical community to help overcome these challenges since they do not impair the quality of the medical image, which is critical for diagnosis, while reducing the size (typically in a 2:1 ratio) of the original image. TIFF format is incredibly adaptable and dynamic, especially in terms of image storage. In contrast to other picture formats, many photos can be saved in a single file. It

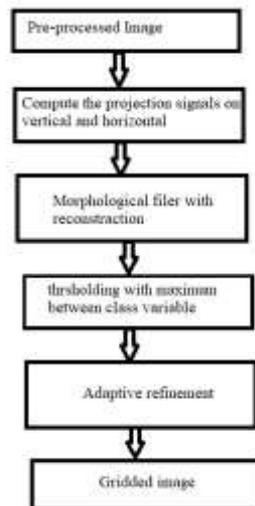
is available in both lossy and lossless formats and is supported by a number of imaging tools. This can capture halftone picture data of the various pixel intensities, manipulating it as an ideal form of graphic storage, processing, and printing format. They used a lossless wavelet compression approach with the MATLAB version R2012a image processing tool package for this project. Wavelet compression is the type of data compression that is well suited for image compression. Its primary objective is to save image data in as little disc coverage as possible. First, after applying the wavelet compression strategy, the image is subjected to a transform that yields as more coefficients as pixels of their image. It was discovered that the dimensions of each image format (434 x 556) remained unchanged after compression.

Chest Computed Tomography scan (CT scan) images [21] are used for prediction of the covid 19 virus because by using the RT-PCR test sometimes it might gives the false result. AI have provided the driven force for the developing various COVID-19 management models. They have used VGG19 Deep Learning model for the Chest image processing because the VGG 19 provide the best results in medical image processing. This model is trained on the 50746 image to classified into two categories covid-19 and healthy. The over fitting is evident from 20 epochs. They consist of images in sixteen bit greyscale TIFF fomate from the dataset images. The images are rescaled into 224*224 since the considered VGG-19 model requisite images in the form of RGB format which consists of size of 224*224 pixels. Its dropout layer is used to make sure of the over fitting. This model is attaining the 93% of accuracy.

Table 1. Accuracy, sensitivity and specify.

Activation Function.	Accuracy. (%)	Sensitivity. (%)	Specificity. (%)
Soft Max	93.00	93.075	92.980
Sigmoid function	82.00	76.019	94.210
ReLU function	80.00	78.03	82.110

Microarray analysis using machine learning models. The dataset consists of the tiff images with sixteen bit gray scale in depth. They have taken six datasets they are SMD, GEO, BCM, UCSF, DeRisi, SIB. This model is tested against four different algorithms [25]. As image contrast is raised, gridding accuracy is considerably boosted when compared to those without enhancement. In each spot region, k-means clustering and considered the k-means clustering they are considered individually. Because the pre-processing operation modifies the image information, the spot intensity is retrieved from the raw image rather than the pre-processing image. This algorithm comparison is used to construct an automatic contrast enrichment in the method that is implemented in the pre-processing involves to enrich the image quality.



3.2 Image segmentation:

Image segmentation [1] for the medical images. The information that dual stream flow is facilitated with the T-NET architecture inside and outside to the encoder-decoder pair. The database ClinicDB that holds frames from twenty nine colorectal colonoscopy in the videos. This collection contains 612 photos in tiff file with a resolution of 384*288 pixels and their related labels. They employed stochastic gradient descent using an insincere rate of learning to create T-Net. The images are tiling technique based on those 7600 photographs for each of the STARE data leave-one-out trails. The accuracy, which indicates the proportion of properly segmented pixels in relation to entire number of pixels to the precisely annotated mask. The T-Net model has an accuracy of 0.931. This strategy is appropriate for both mobile devices and resource-constrained.

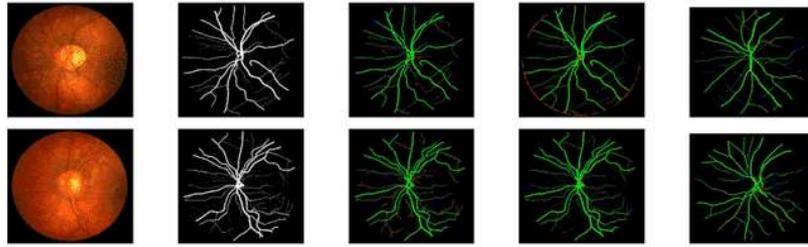


Fig.5 retina image segmentation using U-Net.

The synthetic offline multi-channel [24] pictures were created. In each channel, Each elementary trait in the handwritten trace is represented by a grey level that encodes a single dynamic piece of information relevant to that trait. We also used convolutional neural networks (CNNs) to automatically extract information from offline pictures. The results of three-channel (RGB) and four-channel (TIFF) image production were compared. In the first instance, each elementary trait, velocity, jerk, and pressure applied to produce that trait are encoded utilising three channels. This fourth channel has been added to the first three in the second scenario to encode the acceleration.

The compression of the colour image with “the BMP and TIFF images using the DCT and DWT” [2]. where the DCT is a low level image compression technique that only compresses images with lesser decorative performance as the picture quality metric. DCT only provides a Lossy transform. DWT supports both the Lossy and Lossless transforms. The raw image is taken and transformed into grey scale, and the compression of the grey scale image is exhibited after conversion using the DWT method. They used the 2D wavelet decomposition technique of the approximating the coefficient metrics. When the images are compressed by using this technique it is identified there is not quality and mandatory information loss from the images. The next step in this research will be to create an algorithm that uses DCT or DWT to compress any random image of any resolution or size at a uniform pace without sacrificing image quality.

The image scaling [4] is assessed with the image quality index because images have large range of the experiments like computer vision, image processing and also biomedical commitments. This comparison is based on nearest neighbour, bilinear, and bi cubic interpolation, and It is implemented using the standard picture quality index for various interpolation techniques. This work is being carried out at the PESIT Telecommunications R&D centre. They took 256*256 photos in TIFF, JPEG, and BMP file formats. For scaling factors ranging from 2 to 64, the PSNR value ranges from 70.5 to 60, and the Quality Index ranges from 0.95 to 0.48. In accordance with the universal quality index, the tiff images exceed the outstanding standards of the image quality.

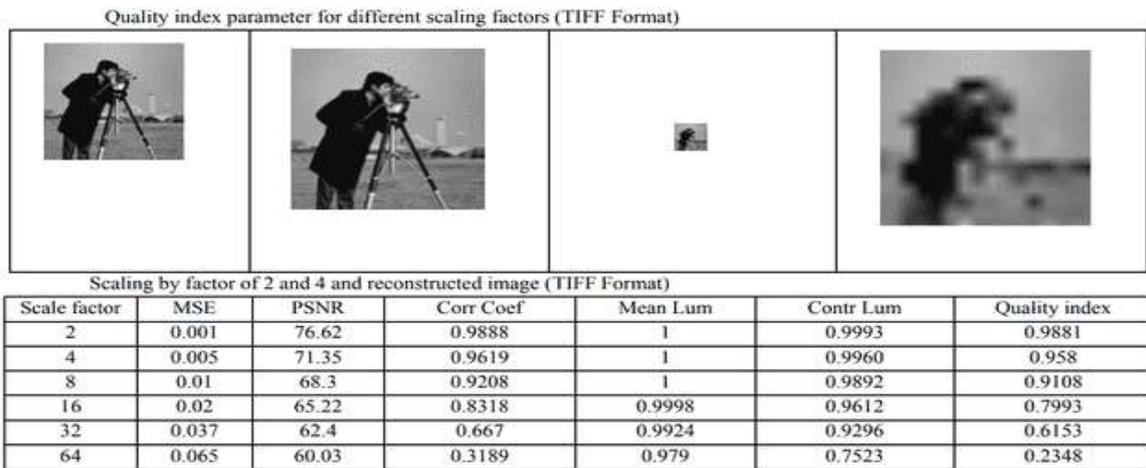


Fig .6 Quality index parameters for different scaling factors (TIFF Format)

Deep learning [7] technique on OCT (optical coherence tomography) information is often used to train the model for classification networks based on 2D images produced from massive data. This study used the following OCT open source datasets, Kermany’s ophthalmology data set and dataset of AIIMS breast tissue, for investigate the consequences of inappropriate dataset segmentation on technique evaluation for two classification model. The dataset consists of the 84,484 TIFF image out of them we have taken the 1000 image from 633 patients for both training and testing .the image are of size 512*496 pixels post the data augmentation. “They have used the 2 layer CNN encoder with one fully connected layer consists of one soft max activation as output layer” [7]. They model products the accuracy as the 96% for the Kermany’s dataset and for the AIIMS dataset it was 98%.

The enhancing plant tissues [11] to capture the microscopic structure which is highly important in growth of the plant tissue. Apart to the tissue cells, they are using XRM(X-ray microscope) in serial imaging of the plant body cells to the predictive and natural stem part should be visualized. The Tagged image file format stack is considered for data integration, visualization, and animation of scan data, as well as exporting picture data. To save the tissue cells in high-resolution format, a 2D 16-bit tag image file format is used. Data from XRM scans were individually segmented using Amira software and a Wacom tablet, combined with the ORS Dragonfly Deep Learning Module, which is free for non-commercial application.

TIFF-Tagged Image File Format [14] is mainly used for bitmap data. TIFF's aim is to used store and exchange source picture data. TIFF file can be as long as 2*32 bytes. TIFF file originates only with an 8-byte image file typically references to an Image File formats. It holds data about the image and links to the image data themselves. Multiband Image Evaluation is the technique used to handle TIFF files in different sectors i.e, Color selection, Crop, Rotate, etc. In this paper, the grey-scaled images are recharging with colored images to get better accuracy to the model. In this paper, they have portrayed the foundation for image processing on the .tiff image file. Using various evaluation techniques like color selection, crop, and rotation, a TIFF image can be thoroughly studied. TIFF is designed to encourage the exchange of digital image information. It can exhibit image data into different color levels as bi-level, grayscale, palette, and wider spaces. A more effective hybrid feature extraction model[22], which was recently proposed, may deliver excellent recognition accuracy while taking up less time during training and classification. In comparison to systems using diagonal-based feature extraction or any other traditional methods of feature extraction, it not only produces higher stage of identification accuracy, though the total time effectiveness has enhanced.

The decrease [15] of image accuracy and quality in comparison to the TIFF format and the various compression settings of the JPEG format analyzed with to the source images obtained by the UAV platform. TIFF is a lossless compression format that is commonly used for photogrammetric and professional photography. Metashape Agisoft was used to create the picture datasets. The processing periods necessary to obtain the 3D models vary depending on the format and compression used. The time processing for the TIFF images is based on different parameters i.e Matching – 7m 46s, Image Orientation – 3m 23s, Depth Mapping- 6m the 20s, and Dense cloud – 23s based on the experimentation results.

The compression of tiff images [16] without data loss by using Artificial neural network techniques. The Multilayer Perception (MLP) methodology, an artificial neural network formulated on the Back-Propagation model for image decompression, is among the solutions. In the survey, they trained a group of (256*256) TIFF images, compressed them using MLP with the various count of neurons in the hidden layer varying from each compression process, and calculated the compression of various accuracy metrics compare the outcomes to generate the source image analysis. This network has three layers input, output, and hidden layer. The same number of neurons (N) must be in both the input and output layers and must connect them fully to the hidden layer. Both ratio of Peak signal-to-noise (PSNR) and MSE (mean square error) they are two factors that affect image quality (MSE). The rebuilt image will be approximately equal to PSNR and progressively proportionate with MSE.

The use of neural networks [20] to segmentation and classification of tissue slides may facilitate clinical diagnosis and contribute to the creation of better diagnostic criteria based on quantitative computational metrics. To enhance network efficiency of prediction, they formed into two-stage segmentation technique. This involves the use of two segmentation networks, of which the first distinguishes hotspot regions at a 1/16th scale and the second among which segments them at the best quality possible. Lossy Compression methods [28] The first step is to modify the data to reduce inter-pixel duplication and accurately group it. Following that, a quantize is utilized to reduce psycho visual redundancy and portray the dense information in as few bits as possible. The quantized bits then intelligently engineered to achieve greater reduces from the redundant coding.

3.3 Character recognition:

Low-cost photo manipulation software make it simple to modify, change, or fabricate digital photographs. The fundamental objective of this research is to discriminate between such a resembled TIFF image and also an source TIFF image. The ratio of likelihood test [LRT] is used to identify resampling data of images in addition to specify the actual images.. The processing is done in 4 steps first the real images are processed then we raw data from it then post processing(white balancing, gamma correction) finally we get original tiff image. We develop a statistical pointer based on the hypothesis testing framework. The concentrate is on creating a detector for JPEG resembled images. Although analyzing probability distributions of pixels in JPEG in the spatial domain is dominating, study in the DCT provine becomes much more conceivable.

Arabic handwriting [23] recognition was introduced. A novel database is created for the gathering, keeping, and retrieving of Arabic text written by hand (AHDB). It is more advanced both in terms of the database's size and the quantity the various authors participating. Additionally, we created an creative, straightforward, and effective in-place tagging the database's operation. It makes it simple for us to extract the word bitmaps. Additionally, we created a pre-processing class contains some practical information pre-processing procedures. The most popular ideas in this publication First-time word identification in Arabic script, utilising a related programme.

An off-line system for recognizing hand printed [13] or machine printed characters. The text stored in a tiff data file from the scanned input, and the output is a sequence of codes signifying the text file's characters. The characters are developed using the Cyrillic character recognition system. The system was developed in 4 stages: architecture, model training, classification and cross validation. The structure of potential memory issues associated with huge TIFF files seriously influenced the approach adopted for line segmentation. The image is beforehand processed by skimming the lines and gathering details about the position and height of such text lines. The filter in testing mode employs the created densely integrated feed forward 3-layer neural network.

Table 2. Comparison table:

Referen ce no.	Limitations	Advantage	Performance metrics	Gaps
1.	Failed to capture information at finer scale and large pooling kernel.	T-Net based convolution neural network is used for medical image segmentation	Dice coefficient is 0.8269 and accuracy is 0.9697	To train the HRF dataset with ERFNet due to large memory footprint and higher resolutions.
2.	The DWT and invert DWT are performed on	Compression without loss of data and quality of the image.	Mean square error :6.88 Peak signal to Noise ratio :30.29	Random image of any resolution or size can be

	the given dataset only.			compressed without degrading its quality.
3.	The genomic and proteomics levels various with patients viruses stage.	cDNA array screening demonstration to identify the generic and metabolic infections.	False discovery rate is 4% and fold change is 6.66	To eliminate the potential using purified subset cell for the initialization of bias into analysis.
4.	MSE and PSNR will not give a better indication for measuring the quality of the image in reconstruction the scaled image.	Universal quality index based image quality indication.	Mean square error is 0.004 And correlation coefficient is 0.9716.	Extension experiment is conducted on all file formats and several interpolation methods.
5.	The automation fails to reject the real time channels that doesn't allocate the requested band.	Transition and rebuilt an imprecise image used algorithm on the TIFF image to as the lossless technique	Mean absolute error is 5% Pixel number quality is 75%	The lost information during compression inorder to reframe the imprecise image.
6.	They does not used any data augmentation techniques and k-fold cross-validation models.	The transfer learning is used efficiently on training the geographical dataset.	Specificity is 0.988 and sensitivity is 0.945	To train the various geographic regions using the deep learning techniques and use hyper parameters tuning to optimize the model.
7.	Discovery and inclusion for growth truth because of missing the information.	optical coherence tomography (OCT) images classification using 2D image from volumetric data.	Accuracy is 97.012 and f1-score is 92%	3-D OCT image classification and optimizing the hyper parameters
8.	medical image analysis of dermatomyositis an autoimmune disease that has not been studied in much detail.	Auto encoder post-processing architecture (APP) is used to increase the performance of the model.	Accuracy is 0.933 and f1 score is 0.8463	By using the same dataset and propose innovative techniques and architectures to improve performance.
9.	This methods is handles a high rate of computational time and memory space	The main advantage is this approach is the supervised methods will give better accuracy.	Dice Score (Dice) is 91.91, , F1-score is 92.12 , mean of Intersection over Union is 85.20 and Recall is 92.82	By using the same model reduce the computation time and memory and achieve high performance.
10.	Efficient Net v2-M model performs well during training but not in the validation.	Transfer learning models are used to increase the effectiveness of the model.	Accuracy is 82.20, Sensitivity is 81.40 Specificity and 94.48	To reduce the losing of the information due to high dropout.
11.	Both for training and validation in machine learning models they used the MR images.	Magnetic resonance imaging (MRI) Support the decisional operation it observes the one step forward.	Area under curve (AUC) is 0.786 and using SVM algorithm it is 0.705	. To collect the blood flow images using computed topographic and validate the robustness.
12.	Machine learning methods are sensitive to overfitting due to their great deal of flexibility.	Among most logistic regression models for estimating a stroke patient's radiological or clinical results.	modified Rankin Score (mRS). using random forest is 0.799 and Neural network is 0.77	Because the selection of parameters is based purely on the literature and expert analysis, predictive patterns in the data could be ignored.
13.	Minimizing the affect of breathing and heartbeat is problematic, especially when the integration time is prolonged.	The vascular quantization method relies on our algorithm can efficiently diagnose ischemic stroke.	Accuracy is 0.879 and F1 score is 0.842	We will evaluate the related respiratory and heartbeat registration methodologies in the selection step.
14.	The training curve is unpredictable and the cross entropy loss training process achieves slowly..	To specifically identify cerebral ischemic lesions in as well CT images, a multi-scale U-Net construction model was developed.	Specificity is 0.89, accuracy is 0.94 and positive prediction score is 0.93	Collaboration among different hospitals and centers is essential and will play a significant role in future work.

15.	In the rebuilding of the temple base and the adjoining topography, it is clear that the two dense point clouds are spaced that is not more than 6 cm.	This study examines how the image's accuracy metrics and quality have degraded or diminished.	Root mean square error is 10.97 and average spectral error is 7.937	As a follow-up to the experiment, future studies will use the normalized time for registration and dense point cloud production that was achieved with the photos in DNG format.
16	In this analysis they have used the online data set they haven't used the raw dataset.	Achieved higher accuracy by using the VGG19 model and used the 500 epochs for the better performance.	Accuracy is 93.08% and sensitivity is 93.73	The CT scan images are used for the COVID-19 and extend the project to extract the critical information.
17	Due to a lack of labelled data, over fitting and improper labelling may result.	Using a triple-branch DSN model, the multi-plane fusion network (MPFN) can extract various plane features and perform well.	Dice score is 72.08 and mean DSC is 71.4	Bring attention to parts of an input image that can be used to segment lesions.
18	They does not able to differentiate the factor of HCMV and serum growth	By using this study we are able ti identify the density of the arrested cell And serum growth.	Accuracy is 91.90 and precision is 89.93	A greater insight of the molecular processes underlying the HCMV life cycle may result in the development of novel therapeutic strategies.
19	This study does not identify the localizing the forgery in the images.	It gives the high performance because of CNN model, CNN can handle larger data sets very easily.	Precision is 93.19 ,Area under curve is 0.9334 and sensitivity is 94.86	In extension to there study they want to compare the different image forgery detection and also investigate on them
20	The compression methods generate the smaller file size but sometimes it loss the necessity of the application.	By using the lossy methods to achieve high accuracy and error rate because of compression file recognized lossless methods.	Mean square error is 5.48 and accuracy is 89.23	In continuation to this compress the uncompressed files like video and audio files without loss of perceptive loss of quality.
21	The training only needs a small amount of parameter tuning.	Isolated incidents are reduced utilizing to prediction dropout layers.	Sensitivity is 0.92, dice similarity coefficient is 0.80 and positive predictive value is 0.87	It does not explain about the deep layers in the fully convolution neural networks.
22	One of the drawbacks of the MR CLEAN Registry is that most of the original, unrestricted images were discarded, retaining only the removed ones.	Deep learning techniques can help with phase classification and perfusion segmentation in this work.	Sensitivity is 0.788 and F1-score is 0.728	In future we implement a AI Model to identify the stroke part.
23	It take more iterations because of remove the selection pressure, it will take more time.	When compared to parametric methods, the stochastic character of the method boosts the likelihood that it will find the global optimal solution and prevents it from becoming stuck in a local optimum.	F1-Score is 0.984 and accuracy is 0.8182	Because small collections of data can still have inconsequential splits, it does not offer the required subtlety we need.
24	The clustering lesion zones in k-means clustering will be larger than the actual lesions. The clustered lesion patches will gradually shrink as K rises..	It has an extremely high precision of 0.880, which really is crucial for avoiding misinterpretation in a clinical setting.	Sensitivity is 0.88 and Dice coefficient is 0.642	.Clinically, 85% of all AIS patients had had a LI stroke. But, in actual practise, it is far challenging to make a diagnosis, particularly when it is too little to be observed.
25	They does not allow the selection of the	Deep learning approach for the feature selection to	Accuracy of rReNet model is 72.60 and VGG19 model is	In continuation they want to explore the dynamic and

	discriminate feature so that classification obtains poor quality.	classify the alzheimer's disease diagnosis through handwriting analysis.	73.05	morphological features by including the in-air features.
--	---	--	-------	--

Feature scope for Research:

The mentioned literatures are mainly concentrated regarding the extracting of features of images using various methods and trained the images by using several techniques such as T-net based convolution based network, image segmentation, cDNA array screening, transfer learning models, auto encoders, U-net, VGG-19, CNN. The respective models will have their disadvantages in respective domains.

- Some of them are does not allow the selection of the discriminate feature so that classification obtains poor quality in the papers [1][3][7][18][25].
- They consider images are in the datasets are mainly in the format of MRI. One of the drawbacks of the MR CLEAN Registry is that most of the original, unrestricted images were discarded, retaining only the removed ones.[3][4][6][10].
- The features obtaining from the images are insufficient such that the performance of the model decreases.[2][5][7][8][12][13][16].
- From the above limitations of several literatures, it can be observed that the TIFF images will contains high resolution features. Hence, it is very tedious task to maintain the prominent features from the images. So, to handle this high quality images techniques such as image segmentation, compression without losing quality and tiling techniques are useful to scale down image size.
- Several feature extraction techniques like HOG, surf, ORB can be applied to the resized images.

3. Conclusion

In this paper the survey mainly focus on the various pre-processing techniques and methodologies that are performed on the various datasets. The tagged image file format is mainly used for the handling of the large size of images. In this study the analysis is about medical images, image segmentation and optimal character reorganization. To perform the models on tiff image the pre-processing is the major task. So to acquire the higher accuracies and feature extraction the tiff images are most preferable. The image tiling for pre-processing, UNet, CNN and Machine learning models are mainly used to handle the tiff images. By this study we can conclude the tiff images are used for the lossless compression, high quality images and best for lossy feature extraction.

References

- [1]. Zhang, R., Zhao, L., Lou, W., Abrigo, J. M., Mok, V. C., Chu, W. C., ... & Shi, L. (2018). Automatic segmentation of acute ischemic stroke from DWI using 3-D fully convolutional DenseNets. *IEEE transactions on medical imaging*, 37(9), 2149-2160.
- [2]. C. -L. Chin et al., "An automated early ischemic stroke detection system using CNN deep learning algorithm," 2017 IEEE 8th International Conference on Awareness Science and Technology (iCAST), 2017, pp. 368-372, doi: 10.1109/ICAwST.2017.8256481.
- [3]. K. C. Ho, W. Speier, H. Zhang, F. Scalzo, S. El-Saden and C. W. Arnold, "A Machine Learning Approach for Classifying Ischemic Stroke Onset Time From Imaging," in *IEEE Transactions on Medical Imaging*, vol. 38, no. 7, pp. 1666-1676, July 2019, doi: 10.1109/TMI.2019.2901445.
- [4]. Kamal, H., Lopez, V., & Sheth, S. A. (2018). Machine learning in acute ischemic stroke neuroimaging. *Frontiers in neurology*, 9, 945.
- [5]. Kim, C., Zhu, V., Obeid, J., & Lenert, L. (2019). Natural language processing and machine learning algorithm to identify brain MRI reports with acute ischemic stroke. *PLoS one*, 14(2), e0212778.
- [6]. Nishio, M., Koyasu, S., Noguchi, S., Kiguchi, T., Nakatsu, K., Akasaka, T., ... & Itoh, K. (2020). Automatic detection of acute ischemic stroke using non-contrast computed tomography and two-stage deep learning model. *Computer Methods and Programs in Biomedicine*, 196, 105711.
- [7]. Van Os, H. J., Ramos, L. A., Hilbert, A., Van Leeuwen, M., Van Walderveen, M. A., Kruijff, N. D., ... & Mr Clean Registry Investigators. (2018). Predicting outcome of endovascular treatment for acute ischemic stroke: potential value of machine learning algorithms. *Frontiers in neurology*, 9, 784..
- [8]. Ho, K. C., Speier, W., Zhang, H., Scalzo, F., El-Saden, S., & Arnold, C. W. (2019). A machine learning approach for classifying ischemic stroke onset time from imaging. *IEEE transactions on medical imaging*, 38(7), 1666-1676.
- [9]. M. Monteiro *et al.*, "Using Machine Learning to Improve the Prediction of Functional Outcome in Ischemic Stroke Patients," in *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 15, no. 6, pp. 1953-1959, 1 Nov.-Dec. 2018, doi: 10.1109/TCBB.2018.2811471.
- [10]. R. Su *et al.*, "autoTICI: Automatic Brain Tissue Reperfusion Scoring on 2D DSA Images of Acute Ischemic Stroke Patients," in *IEEE Transactions on Medical Imaging*, vol. 40, no. 9, pp. 2380-2391, Sept. 2021, doi: 10.1109/TMI.2021.3077113.
- [11]. L. Tomasetti, L. J. Hølllesli, K. Engan, K. D. Kurz, M. W. Kurz and M. Khanmohammadi, "Machine Learning Algorithms Versus Thresholding

- to Segment Ischemic Regions in Patients With Acute Ischemic Stroke," in *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 2, pp. 660-672, Feb. 2022, doi: 10.1109/JBHI.2021.3097591.
- [12]. V. L. Coli *et al.*, "Detection of Brain Strokes Using Microwave Tomography," 2018 IEEE International Symposium on Antennas and Propagation & USNC/URSI National Radio Science Meeting, 2018, pp. 223-224, doi: 10.1109/APUSNCURSINRSM.2018.8609404.
- [13]. Pathanjali C , Monisha G , Priya T , Ruchita Sudarshan K, Samyuktha Bhaskar, 2020, Machine Learning for Predicting Ischemic Stroke, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 09, Issue 05 (May 2020).
- [14]. Li S, Zheng J, Li D. Precise segmentation of non-enhanced computed tomography in patients with ischemic stroke based on multi-scale U-Net deep network model. *Comput Methods Programs Biomed.* 2021 Sep;208:106278. doi: 10.1016/j.cmpb.2021.106278. Epub 2021 Jul 9. PMID: 34274610.
- [15]. Yifan Hao, Aiyan Ji, Rongrong Xing, Wenqing Zhu, Baohong Jiang, Yi Jian, Hao
- [16]. Chen, Capillaries segmentation of NIR-II images and its application in ischemic stroke, *Computers in Biology and Medicine*, Volume 147,2022,105742, ISSN 0010-4825.
- [17]. To, M.N.N., Kim, H.J., Roh, H.G. et al. Deep regression neural networks for collateral imaging from dynamic susceptibility contrast-enhanced magnetic resonance perfusion in acute ischemic stroke. *Int J CARS* 15, 151–162 (2020). <https://doi.org/10.1007/s11548-019-02060-7>
- [18]. Amish Kumar, Neha Upadhyay, Palash Ghosal, Tamal Chowdhury, Dipayan Das, Amritendu Mukherjee, Debashis Nandi, CSNet: A new DeepNet framework for ischemic stroke lesion segmentation, *Computer Methods and Programs in Biomedicine*, Volume 193,2020,105524,ISSN 0169-2607.
- [19]. L. Zhang et al., "Ischemic Stroke Lesion Segmentation Using Multi-Plane Information Fusion," in *IEEE Access*, vol. 8, pp. 45715-45725, 2020, doi: 10.1109/ACCESS.2020.2977415.
- [20]. Bin Zhao, Zhiyang Liu, Guohua Liu, Chen Cao, Song Jin, Hong Wu, Shuxue Ding, "Deep Learning-Based Acute Ischemic Stroke Lesion Segmentation Method on Multimodal MR Images Using a Few Fully Labeled Subjects", *Computational and Mathematical Methods in Medicine*, vol. 2021, Article ID 3628179, 13 pages, 2021. <https://doi.org/10.1155/2021/3628179>
- [21]. Samak, Z. A., Clatworthy, P., & Mirmehdi, M. (2022). FeMA: Feature matching auto-encoder for predicting ischaemic stroke evolution and treatment outcome. *Computerized Medical Imaging and Graphics*, 99, 102089.
- [22]. Zhao, B., Liu, Z., Liu, G., Cao, C., Jin, S., Wu, H., & Ding, S. (2021). Deep learning-based acute ischemic stroke lesion segmentation method on multimodal MR images using a few fully labeled subjects. *Computational and Mathematical Methods in Medicine*, 2021.
- [23]. Platscher, M., Zopes, J., & Federau, C. (2022). Image translation for medical image generation: Ischemic stroke lesion segmentation. *Biomedical Signal Processing and Control*, 72, 103283.
- [24]. Islam, M., Vaidyanathan, N. R., Jose, V., & Ren, H. (2018, September). Ischemic stroke lesion segmentation using adversarial learning. In *International MICCAI Brainlesion Workshop* (pp. 292-300). Springer, Cham.
- [25]. Kappelhof, N., Ramos, L. A., Kappelhof, M., van Os, H. J., Chalos, V., van Kranendonk, K. R., ... & Marquering, H. A. (2021). Evolutionary algorithms and decision trees for predicting poor outcome after endovascular treatment for acute ischemic stroke. *Computers in Biology and Medicine*, 133, 104414.