***Application and Construction of Deep Learning Networks in Medical Imaging***

**ABSTRACT**

* Breast cancer continues to be among the leading causes of death for women and much effort has been expended in the form of screening programs for prevention.
* Given the exponential growth in the number of mammograms collected by these programs, computer-assisted diagnosis has become a necessity.
* Computer-assisted detection techniques developed to date to improve diagnosis without multiple systematic readings have not resulted in a significant improvement in performance measures.
* In this context, the use of automatic image processing techniques resulting from deep learning represents a promising avenue for assisting in the diagnosis of breast cancer.

**Objective**

* In this paper, we present a deep learning approach based on a Convolutional Neural Network (CNN) and RNN model for multi-class breast cancer classification.
* The proposed approach aims to classify the breast tumors as benign or malignant.

**INTRODUCTION**

**Breast cancer**

Breast cancer is a malignant tumor in or around breast tissue. It usually begins as a lump or calcium deposit that develops from abnormal cell growth. Most breast lumps are benign, but some can be pre-cancerous or cancerous. Breast cancer can be localized (initially appearing within the breast) or metastatic (spread to another part of the body).

Your doctor will likely perform a physical exam to evaluate a breast lump. To determine whether that lump is benign, your doctor may order mammography, breast ultrasound, breast MRI, PET/CT or scintimammography. If the lump is indeed benign, no further action may be needed. However, your doctor may want to monitor it to see if it changes, grows, or disappears over time. If the tests are inconclusive, your doctor may perform a biopsy using ultrasound, x-ray, or magnetic resonance imaging guidance. Breast cancer treatment depends on the tumor's size, extent of disease spread, type, receptor status, tumor growth rate and the patient’s general health. Treatments include surgery, radiation therapy, chemotherapy, hormone therapy or a combination thereof.

**Benign Breast Disease**

Benign breast disease makes you more prone to getting breast lumps. Finding a lump can be scary, but these breast changes are benign (not cancer). Certain types of breast disease increase your risk of breast cancer. You should notify your healthcare provider about any breast lumps or changes. Most noncancerous lumps go away without treatment.

**Mitotic Cells**

As mentioned above, the number of mitotic cells in breast histopathology images is one of the three components (the other two being tubule formation, nuclear pleomorphism) for grading of breast cancer tissue slides. Automatic detection and counting of mitotic cells is a challenging task since the biological variability of the mitotic cells makes their detection extremely difficult . The importance of studying the mitotic count arises from the fact that it evaluat es the aggressiveness of the tumor. Mitotic c ells have the ability to replicate each of their chromosomes and separate the chromosomes in their nuclei into two identical sets of chromosomes where each set is in its own new nucleus and ending by having two new cells.

***Here are four phases of mitosis:***

Prophase: Prophase is the first phase in the mitosis process. During this phase, nucleoli disappear, and the chromatin (DNA and associated proteins) can be observed with a light based microscope. The chromosome consist s of two sister chromatids that have the same genetic information. The two sister chromatids are connected together through a centromere. As the centrioles reach opposite ends of the cell, the spindle fibers from each of the centrioles attach to each chromosome at a specialized protein structure called the kinetochore. The kinetochore is located at the centromere of each chromosome. Other spindle fibers elongate, but instead of attaching to chromosomes, they interact with spindle from the opposite pole.

Metaphase: Along the middle of the cell nucleus , the chromosomes will be aligned. The resulted line is called the metaphase plate. This process will help to ensure in next phases that each new nucleus will have a copy of each chromosome when the aligned chromosomes are separated.

Histopathology slides are one of the most applicable resources for pathology studies. As observation of these kinds of slides even by skillful pathologists is a tedious and time-consuming activity, computerizing this procedure aids the experts to have faster analysis with more case studies per day. In this paper, an automatic mitosis detection system (AMDS) for breast cancer histopathological slide images is proposed. In the proposed AMDS, the general phases of an automatic image based analyzer are considered and in each phase, some special innovations are employed. In the pre-processing step to segment the input digital histopathology images more precisely, 2D anisotropic diffusion filters are applied to them. In the training segmentation phase, the histopathological slide images are segmented based on RGB contents of their pixels using maximum likelihood estimation. Then, the mitosis and non-mitosis candidates are processed and hence that their completed local binary patterns are extracted object-wise. For the classification phase, two subsequently non-linear support vector machine classifiers are trained pixel-wise and object-wise, respectively. For the evaluation of the proposed AMDS, some object and region based measures are employed. Having computed the evaluation criteria, our proposed method performs more efficient according to f-measure metric (70.94% for Aperio XT scanner images and 70.11% for Hamamatsu images) than the methods proposed by other participants at Mitos-ICPR2012 contest in breast cancer histopathological images. The experimental results show the higher performance of the proposed AMDS compared with other competitive systems proposed in Mitos-ICPR2012 contest.

**DOMAIN EXPLANATION**

**IMAGE PROCESSING**

* 1. **What is an image?**

An image is an array, or a matrix, of square pixels (picture elements) arranged in columns and rows.

**Figure 1: An image — an array or a matrix of pixels arranged in columns and rows.**

In a (8-bit) greyscale image each picture element has an assigned intensity that ranges from 0 to 255. A grey scale image is what people normally call a black and white image, but the name emphasizes that such an image will also include many shades of grey.

**Figure 2: Each pixel has a value from 0 (black) to 255 (white). The possible range of the pixel values depend on the colour depth of the image, here 8 bit = 256 tones or greyscales.**

A normal grayscale image has 8 bit colour depth = 256 grayscales. A “true colour” image has 24 bit colour depth = 8 x 8 x 8 bits = 256 x 256 x 256 colours = ~16 million colours.

**Figure 3: A true-colour image assembled from three grayscale images coloured red, green and blue. Such an image may contain up to 16 million different colours.**

Some grayscale images have more grayscales, for instance 16 bit = 65536 grayscales. In principle three grayscale images can be combined to form an image with 281,474,976,710,656 grayscales.

There are two general groups of ‘images’: vector graphics (or line art) and bitmaps (pixel-based or ‘images’). Some of the most common file formats are:

GIF — an 8-bit (256 colour), non-destructively compressed bitmap format. Mostly used for web. Has several sub-standards one of which is the animated GIF.

JPEG — a very efficient (i.e. much information per byte) destructively compressed 24 bit (16 million colours) bitmap format. Widely used, especially for web and Internet (bandwidth-limited).

TIFF — the standard 24 bit publication bitmap format. Compresses non-destructively with, for instance, Lempel-Ziv-Welch (LZW) compression.

PS — Postscript, a standard vector format. Has numerous sub-standards and can be difficult to transport across platforms and operating systems.

PSD – a dedicated Photoshop format that keeps all the information in an image including all the layers.

Pictures are the most common and convenient means of conveying or transmitting information. A picture is worth a thousand words. Pictures concisely convey information about positions, sizes and inter relationships between objects. They portray spatial information that we can recognize as objects. Human beings are good at deriving information from such images, because of our innate visual and mental abilities. About 75% of the information received by human is in pictorial form. An image is digitized to convert it to a form which can be stored in a computer's memory or on some form of storage media such as a hard disk or CD-ROM. This digitization procedure can be done by a scanner, or by a video camera connected to a frame grabber board in a computer. Once the image has been digitized, it can be operated upon by various image processing operations.

Image processing operations can be roughly divided into three major categories, Image Compression, Image Enhancement and Restoration, and Measurement Extraction. It involves reducing the amount of memory needed to store a digital image. Image defects which could be caused by the digitization process or by faults in the imaging set-up (for example, bad lighting) can be corrected using Image Enhancement techniques. Once the image is in good condition, the Measurement Extraction operations can be used to obtain useful information from the image.

Some examples of Image Enhancement and Measurement Extraction are given below. The examples shown all operate on 256 grey-scale images. This means that each pixel in the image is stored as a number between 0 to 255, where 0 represents a black pixel, 255 represents a white pixel and values in-between represent shades of grey. These operations can be extended to operate on colour images. The examples below represent only a few of the many techniques available for operating on images. Details about the inner workings of the operations have not been given, but some references to books containing this information are given at the end for the interested reader.

**Images and pictures**

As we mentioned in the preface, human beings are predominantly visual creatures: we rely heavily on our vision to make sense of the world around us. We not only look at things to identify and classify them, but we can scan for differences, and obtain an overall rough feeling for a scene with a quick glance. Humans have evolved very precise visual skills: we can identify a face in an instant; we can differentiate colors; we can process a large amount of visual information very quickly.

However, the world is in constant motion: stare at something for long enough and it will change in some way. Even a large solid structure, like a building or a mountain, will change its appearance depending on the time of day (day or night); amount of sunlight (clear or cloudy), or various shadows falling upon it. We are concerned with single images: snapshots, if you like, of a visual scene. Although image processing can deal with changing scenes, we shall not discuss it in any detail in this text. For our purposes, an image is a single picture which represents something. It may be a picture of a person, of people or animals, or of an outdoor scene, or a microphotograph of an electronic component, or the result of medical imaging. Even if the picture is not immediately recognizable, it will not be just a random blur.

Image processing involves changing the nature of an image in order to either

1. Improve its pictorial information for human interpretation,

2. Render it more suitable for autonomous machine perception.

We shall be concerned with digital image processing, which involves using a computer to change the nature of a digital image. It is necessary to realize that these two aspects represent two separate but equally important aspects of image processing. A procedure which satisfies condition, a procedure which makes an image look better may be the very worst procedure for satisfying condition. Humans like their images to be sharp, clear and detailed; machines prefer their images to be simple and uncluttered.

**Images and digital images**

Suppose we take an image, a photo, say. For the moment, lets make things easy and suppose the photo is black and white (that is, lots of shades of grey), so no colour. We may consider this image as being a two dimensional function, where the function values give the brightness of the image at any given point. We may assume that in such an image brightness values can be any real numbers in the range (black) (white).

A digital image from a photo in that the values are all discrete. Usually they take on only integer values. The brightness values also ranging from 0 (black) to 255 (white). A digital image can be considered as a large array of discrete dots, each of which has a brightness associated with it. These dots are called picture elements, or more simply pixels. The pixels surrounding a given pixel constitute its neighborhood. A neighborhood can be characterized by its shape in the same way as a matrix: we can speak of a neighborhood,. Except in very special circumstances, neighborhoods have odd numbers of rows and columns; this ensures that the current pixel is in the centre of the neighborhood.

**Image Processing Fundamentals:**

**Pixel:**

In order for any digital computer processing to be carried out on an image, it must first be stored within the computer in a suitable form that can be manipulated by a computer program. The most practical way of doing this is to divide the image up into a collection of discrete (and usually small) cells, which are known as pixels. Most commonly, the image is divided up into a rectangular grid of pixels, so that each pixel is itself a small rectangle. Once this has been done, each pixel is given a pixel value that represents the color of that pixel.

It is assumed that the whole pixel is the same color, and so any color variation that did exist within the area of the pixel before the image was discretized is lost. However, if the area of each pixel is very small, then the discrete nature of the image is often not visible to the human eye.

Other pixel shapes and formations can be used, most notably the hexagonal grid, in which each pixel is a small hexagon. This has some advantages in image processing, including the fact that pixel connectivity is less ambiguously defined than with a square grid, but hexagonal grids are not widely used. Part of the reason is that many image capture systems (e.g. most CCD cameras and scanners) intrinsically discretize the captured image into a rectangular grid in the first instance.

**Pixel Connectivity**

The notation of pixel connectivity describes a relation between two or more pixels. For two pixels to be connected they have to fulfill certain conditions on the pixel brightness and spatial adjacency.

First, in order for two pixels to be considered connected, their pixel values must both be from the same set of values V. For a grayscale image, V might be any range of graylevels, e.g. V={22,23,...40}, for a binary image we simple have V={1}.

To formulate the adjacency criterion for connectivity, we first introduce the notation of neighborhood. For a pixel p with the coordinates (x,y) the set of pixels given by:

Eqn:eqnla1

is called its 4-neighbors. Its 8-neighbors are defined as

Eqn:eqnla2

From this we can infer the definition for 4- and 8-connectivity:

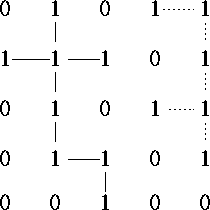
Two pixels p and q, both having values from a set V are 4-connected if q is from the set Eqn:eqnlan4and 8-connected if q is from Eqn:eqnlan8.

General connectivity can either be based on 4- or 8-connectivity; for the following discussion we use 4-connectivity.

A pixel p is connected to a pixel q if p is 4-connected to q or if p is 4-connected to a third pixel which itself is connected to q. Or, in other words, two pixels q and p are connected if there is a path from p and q on which each pixel is 4-connected to the next one.

A set of pixels in an image which are all connected to each other is called a connected component. Finding all connected components in an image and marking each of them with a distinctive label is called connected component labeling.

An example of a binary image with two connected components which are based on 4-connectivity can be seen in Figure 1. If the connectivity were based on 8-neighbors, the two connected components would merge into one.



**Figure 1** Two connected components based on 4-connectivity.

# Pixel Values

Each of the pixels that represents an image stored inside a computer has a pixel value which describes how bright that pixel is, and/or what color it should be. In the simplest case of binary images, the pixel value is a 1-bit number indicating either foreground or background. For a gray scale images, the pixel value is a single number that represents the brightness of the pixel. The most common motepixel format is the byte image, where this number is stored as an 8-bit integer giving a range of possible values from 0 to 255. Typically zero is taken to be black, and 255 is taken to be white. Values in between make up the different shades of gray.

To represent colour images, separate red, green and blue components must be specified for each pixel (assuming an RGB colour space), and so the pixel `value' is actually a vector of three numbers. Often the three different components are stored as three separate `grayscale' images known as color planes (one for each of red, green and blue), which have to be recombined when displaying or processing. Multispectral Images can contain even more than three components for each pixel, and by extension these are stored in the same kind of way, as a vector pixel value, or as separate color planes.

The actual grayscale or color component intensities for each pixel may not actually be stored explicitly. Often, all that is stored for each pixel is an index into a colour map in which the actual intensity or colors can be looked up. Although simple 8-bit integers or vectors of 8-bit integers are the most common sorts of pixel values used, some image formats support different types of value, for instance 32-bit signed integers or floating point values. Such values are extremely useful in image processing as they allow processing to be carried out on the image where the resulting pixel values are not necessarily 8-bit integers. If this approach is used then it is usually necessary to set up a colormap which relates particular ranges of pixel values to particular displayed colors.

**Pixels, with a neighborhood:**

****

**Color scale**

The two main color spaces are **RGB** and **CMYK.**

**RGB**

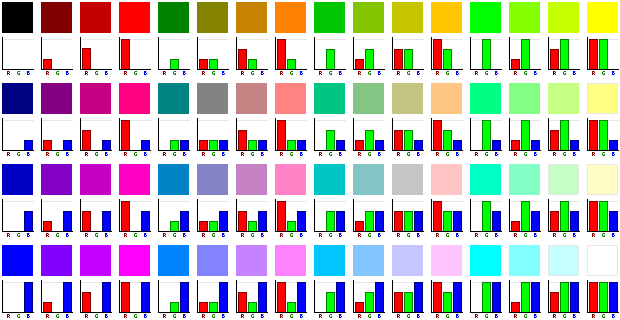
The **RGB color model** is an additive color model in which red, green, and blue light are added together in various ways to reproduce a broad array of colors. RGB uses additive color mixing and is the basic color model used in television or any other medium that projects color with light. It is the basic color model used in computers and for web graphics, but it cannot be used for print production.

The secondary colors of RGB – cyan, magenta, and yellow – are formed by mixing two of the primary colors (**red, green** or **blue**) and excluding the third color. Red and green combine to make yellow, green and blue to make cyan, and blue and red form magenta. The combination of red, green, and blue in full intensity makes white.



The additive model of RGB. Red, green, and blue are the primary stimuli for human color perception and are the primary additive colours.

To see how different RGB components combine together, here is a selected repertoire of colors and their respective relative intensities for each of the red, green, and blue components:

[](http://en.wikipedia.org/wiki/Image:RGB-combinations.png)

**\*Typical uses of MATLAB include:-**

- Math and computation.

-Algorithm development

-Data acquisition

-Modeling, simulation, and prototyping

-Data analysis, exploration, and visualization

-Scientific and engineering graphics

-Application development, including graphical user interface building

**Some applications:**

Image processing has an enormous range of applications; almost every area of science and technology can make use of image processing methods. Here is a short list just to give some indication of the range of image processing applications.

1. Medicine

* Inspection and interpretation of images obtained from X-rays, MRI or CAT scans,
* Analysis of cell images, of chromosome karyotypes.

2. Agriculture

* Satellite/aerial views of land, for example to determine how much land is being used for different purposes, or to investigate the suitability of different regions for different crops,
* Inspection of fruit and vegetables distinguishing good and fresh produce from old.

3. Industry

* Automatic inspection of items on a production line,
* Inspection of paper samples.

4. Law enforcement

* Fingerprint analysis,
* Sharpening or de-blurring of speed-camera images.

**Aspects of image processing:**

It is convenient to subdivide different image processing algorithms into broad subclasses. There are different algorithms for different tasks and problems, and often we would like to distinguish the nature of the task at hand.

* **Image enhancement**: This refers to processing an image so that the result is more suitable for a particular application.

Example include:

* sharpening or de-blurring an out of focus image,
* highlighting edges,
* improving image contrast, or brightening an image,
* Removing noise.
* **Image restoration**. This may be considered as reversing the damage done to an image by a known cause, for example:
* removing of blur caused by linear motion,
* removal of optical distortions,
* Removing periodic interference.
* **Image segmentation**. This involves subdividing an image into constituent parts, or isolating certain aspects of an image:
* circles, or particular shapes in an image,
* In an aerial photograph, identifying cars, trees, buildings, or roads.

These classes are not disjoint; a given algorithm may be used for both image enhancement or for image restoration. However, we should be able to decide what it is that we are trying to do with our image: simply make it look better (enhancement), or removing damage (restoration).

**An image processing task**

We will look in some detail at a particular real-world task, and see how the above classes may be used to describe the various stages in performing this task. The job is to obtain, by an automatic process, the postcodes from envelopes. Here is how this may be accomplished:

* **Acquiring the image**: First we need to produce a digital image from a paper envelope. This can be done using either a CCD camera, or a scanner.
* **Preprocessing:** This is the step taken before the major image processing task. The problem here is to perform some basic tasks in order to render the resulting image more suitable for the job to follow. In this case it may involve enhancing the contrast, removing noise, or identifying regions likely to contain the postcode.
* **Segmentation**: Here is where we actually get the postcode; in other words we extract from the image that part of it which contains just the postcode.
* **Representation and description** These terms refer to extracting the particular features which allow us to differentiate between objects. Here we will be looking for curves, holes and corners which allow us to distinguish the different digits which constitute a postcode.
* **Recognition and interpretation**: This means assigning labels to objects based on their descriptors (from the previous step), and assigning meanings to those labels. So we identify particular digits, and we interpret a string of four digits at the end of the address as the postcode.

**EXISTING SYSTEM**

Medical image segmentation is considered as a hot research topic. Several researchers have suggested various methodologies and algorithms for image segmentation. For example, Bandhyopadhya proposed a brain tumor segmentation method based on K-means clustering technique. The method consists of three steps: K-means algorithm based segmentation, local standard deviation guided grid based coarse grain localization, and local standard deviation guided grid based fine grain localization. The extraction of the brain tumor region from the processed image requires the segmentation of the brain MRI images to two segments. One segment contains the normal brain cells consisting of Grey Matter (GM), White Matter (WM), and the Cerebral Spinal Fluid (CSF). The second segment contains the tumor cells of the brain. The segmentation technique is constraint by the fact that the images need to be of adjacent imaging layer. The image fusion method gave a good result in fusing multiple images. In particular cases, it resulted in the loss of intensity. Moreover, it also ignored the finer anatomic details, such as twists and turns in the boundary of the tumor or overlapping region of gray and white matters in the brain. They proposed an approach of Spatial Fuzzy C-means (PET-SFCM) clustering algorithm on Positron Emission Tomography (PET) scan image datasets. The algorithm is joining the spatial neighborhood information with classical FCM and updating the objective function of each cluster. Spatial relationship of neighboring pixel is an aid of image segmentation. These neighboring pixels are highly renovated the same feature data. In spatial domain, the memberships of the neighbor centered are specified to obtain the cluster distribution statistics. They calculated the weighting function based on these statistics and applied into the membership function. Their algorithm is tested on data collection of patients with Alzheimer’s disease. They did not calculate objective based quality assessment that could analyze images and did not report their quality without human involvement. Proposed system that using a convolution neural network (CNN) as pixel classifier for the segmentation process of some X-ray images. The system analyzes each pixel from the image and tries to classify them into two classes: bone and non-bone. They attempted to separate the bone tissue area from the rest of the image. Their CNN obtained the best results in contrast to other configurations. For ensuring a minimum training time of the network, they used only the interest areas from an image. Their method recognized the significant bone areas, but the problems appeared when the bone area presented irregularities and take more execution time in training. Introduced image segmentation using K-means clustering, Expectation Maximization (EM), and Normalized Cuts (NC). They analyzed the two former unsupervised learning algorithms and compared them with a graph-based algorithm, the Normalized Cut algorithm. They applied the partitioning algorithm to gray-scaled images with varying value of k (number of clusters). For smaller values of k, the K-means and EM algorithms give good results. For larger values of k, the segmentation is very coarse; many clusters appear in the images at discrete places. The NCuts algorithm gave good results for larger value of k, but it takes a long time.

Proposed color image segmentation using K-Medoids Clustering. The idea of the algorithm is to find clusters of objects by finding the Medoids for each cluster. Each remaining object is clustered with the Medoid or representative objects to which it is the most similar. K-Medoids method uses representative objects as reference points rather than taking the mean value of the objects in each cluster. The algorithm takes the input parameter k and the number of clusters to be partitioned among a set of n objects. The segmented images are highly dependent on the number of segments or centers. They did not consider finding optimal number of segments to provide more accurate results. Proposed image segmentation technique based on K-means, K-Mediods, and Hierarchical clustering technologies. They made a comparison between these three clustering techniques on natural images to find the advantages and disadvantages of each algorithm. After applying these algorithms, they mentioned that the K-means Clustering method has better performance and easy to implement than other clustering methods. On the other hand, other several researchers have suggested various hybrid algorithms for image segmentation. The integration between K-means and Fuzzy C-means. They chose the number of clusters, fuzziness, distance, and stopping the criterion. Then, they initialized the memberships randomly or getting from K-means and in iterations, recalculating centers and memberships until the objective function reached. The advantage of their method is that it can deal with overlapping grayscale intensities. The disadvantage of their proposed method is that it cannot clearly defined borders between tissues successfully. Although, it minimizes the within-class sum square errors, but its performance degrade when applied to noise corrupted images. They solved this problem by the preprocessing step before applying the integration. They compared their result with KM, FCM, and the integration FKM in case of under-segmentation and over-segmentation. They proved that FKM gives minimum under or over-segmentation, but they did not demonstrate what about time of each algorithm or in the integration method.

**Disadvantages:**

* There is only 27% of the cells are used to detect the cancerous cell.
* There are manual image segmentation process is involved, so that the image segmentation will not accurate.

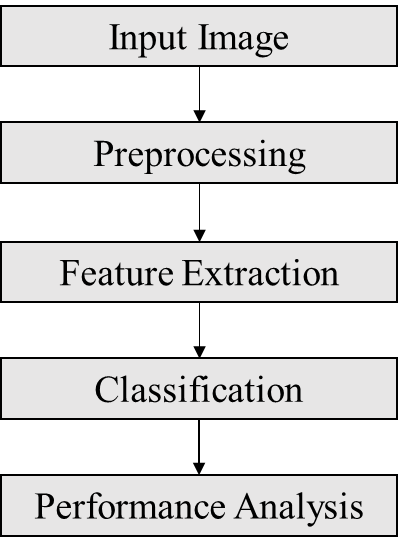
**PROPOSED SYSTEM**

* In this proposed system, the spiral drawing data is taken as an input.
* Then the preprocessing is done by means of the image resize and channel separation.
* Then the segmentation is done to extract the written region.
* Then the features are extracted by means of the texture based. Hence the GLCM algorithm is used to extract the feature values from the image.
* After the feature extraction, the classification is done to recognize the input is normal or abnormal.
* Finally the performance of the classifier is measured by means of the accuracy, sensitivity and specificity.

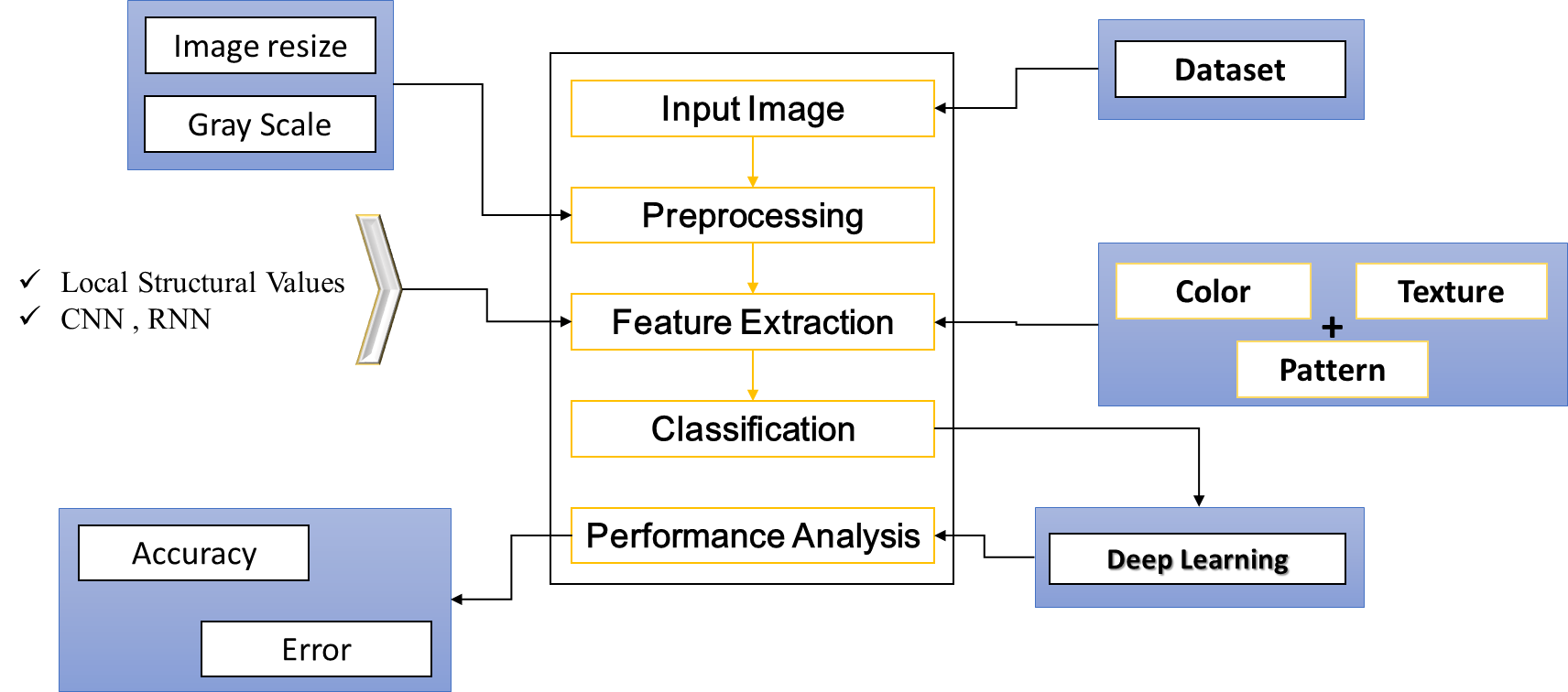
**Advantages:**

* It improve the achieved in feature stability.
* It improve the precision of the classification.

**SYSTEM ARCHITECTURE:**

****

**FLOW DIAGRAM:**

****

**TESTING OF PRODUCT**

**Testing of Product:**

System testing is the stage of implementation, which aimed at ensuring that system works accurately and efficiently before the live operation commence. Testing is the process of executing a program with the intent of finding an error. A good test case is one that has a high probability of finding an error. A successful test is one that answers a yet undiscovered error.

Testing is vital to the success of the system.  System testing makes a logical assumption that if all parts of the system are correct, the goal will be successfully achieved.  The candidate system is subject to variety of tests-on-line response, Volume Street, recovery and security and usability test.  A series of tests are performed before the system is ready for the user acceptance testing.  Any engineered product can be tested in one of the following ways.  Knowing the specified function that a product has been designed to from, test can be conducted to demonstrate each function is fully operational.  Knowing the internal working of a product, tests can be conducted to ensure that “al gears mesh”, that is the internal operation of the product performs according to the specification and all internal components have been adequately exercised.

**UNIT TESTING:**

Unit testing is the testing of each module and the integration of the overall system is done.  Unit testing becomes verification efforts on the smallest unit of software design in the module.  This is also known as ‘module testing’.  The modules of the system are tested separately.  This testing is carried out during the programming itself.  In this testing step, each model is found to be working satisfactorily as regard to the expected output from the module.  There are some validation checks for the fields.  For example, the validation check is done for verifying the data given by the user where both format and validity of the data entered is included.  It is very easy to find error and debug the system.

**INTEGRATION TESTING:**

Data can be lost across an interface, one module can have an adverse effect on the other sub function, when combined, may not produce the desired major function.  Integrated testing is systematic testing that can be done with sample data.  The need for the integrated test is to find the overall system performance. There are two types of integration testing. They are:

1. Top-down integration testing.
2. Bottom-up integration testing.

**WHITE BOX TESTING:**

White Box testing is a test case design method that uses the control structure of the procedural design to drive cases.  Using the white box testing methods, we derived test cases that guarantee that all independent paths within a module have been exercised at least once.

**BLACK BOX TESTING:**

* + Black box testing is done to find incorrect or missing function
  + Interface error
  + Errors in external database access
  + Performance errors
  + Initialization and termination errors

In ‘functional testing’, is performed to validate an application conforms to its specifications of correctly performs all its required functions. So this testing is also called ‘black box testing’.  It tests the external behavior of the system.  Here the engineered product can be tested knowing the specified function that a product has been designed to perform, tests can be conducted to demonstrate that each function is fully operational.

**VALIDATION TESTING:**

After the culmination of black box testing, software is completed assembly as a package, interfacing errors have been uncovered and corrected and final series of software validation tests begin validation testing can be defined as many, but a single definition is that validation succeeds when the software functions in a manner that can be reasonably expected by the customer.

# USER ACCEPTANCE TESTING:

User acceptance of the system is the key factor for the success of the system.  The system under consideration is tested for user acceptance by constantly keeping in touch with prospective system at the time of developing changes whenever required.

# OUTPUT TESTING:

After performing the validation testing, the next step is output asking the user about the format required testing of the proposed system, since no system could be useful if it does not produce the required output in the specific format.  The output displayed or generated by the system under consideration.  Here the output format is considered in two ways.  One is screen and the other is printed format.  The output format on the screen is found to be correct as the format was designed in the system phase according to the user needs.  For the hard copy also output comes out as the specified requirements by the user. Hence the output testing does not result in any connection in the system.

**System Implementation:**

Implementation of software refers to the final installation of the package in its real environment, to the satisfaction of the intended users and the operation of the system. The people are not sure that the software is meant to make their job easier.

* The active user must be aware of the benefits of using the system
* Their confidence in the software built up
* Proper guidance is impaired to the user so that he is comfortable in using the application

Before going ahead and viewing the system, the user must know that for viewing the result, the server program should be running in the server. If the server object is not running on the server, the actual processes will not take place.

User Training:

To achieve the objectives and benefits expected from the proposed system it is essential for the people who will be involved to be confident of their role in the new system. As system becomes more complex, the need for education and training is more and more important.

Education is complementary to training. It brings life to formal training by explaining the background to the resources for them. Education involves creating the right atmosphere and motivating user staff. Education information can make training more interesting and more understandable.

**Training on the Application Software:**

After providing the necessary basic training on the computer awareness, the users will have to be trained on the new application software. This will give the underlying philosophy of the use of the new system such as the screen flow, screen design, type of help on the screen, type of errors while entering the data, the corresponding validation check at each entry and the ways to correct the data entered. This training may be different across different user groups and across different levels of hierarchy.

**Operational Documentation:**

Once the implementation plan is decided, it is essential that the user of the system is made familiar and comfortable with the environment. A documentation providing the whole operations of the system is being developed. Useful tips and guidance is given inside the application itself to the user. The system is developed user friendly so that the user can work the system from the tips given in the application itself.

**System Maintenance:**

The maintenance phase of the software cycle is the time in which software performs useful work. After a system is successfully implemented, it should be maintained in a proper manner. System maintenance is an important aspect in the software development life cycle. The need for system maintenance is to make adaptable to the changes in the system environment. There may be social, technical and other environmental changes, which affect a system which is being implemented. Software product enhancements may involve providing new functional capabilities, improving user displays and mode of interaction, upgrading the performance characteristics of the system. So only thru proper system maintenance procedures, the system can be adapted to cope up with these changes. Software maintenance is of course, far more than “finding mistakes”.

**Corrective Maintenance:**

The first maintenance activity occurs because it is unreasonable to assume that software testing will uncover all latent errors in a large software system. During the use of any large program, errors will occur and be reported to the developer. The process that includes the diagnosis and correction of one or more errors is called Corrective Maintenance.

**Adaptive Maintenance:**

The second activity that contributes to a definition of maintenance occurs because of the rapid change that is encountered in every aspect of computing. Therefore Adaptive maintenance termed as an activity that modifies software to properly interfere with a changing environment is both necessary and commonplace.

**Perceptive Maintenance:**

The third activity that may be applied to a definition of maintenance occurs when a software package is successful. As the software is used, recommendations for new capabilities, modifications to existing functions, and general enhancement are received from users. To satisfy requests in this category, Perceptive maintenance is performed. This activity accounts for the majority of all efforts expended on software maintenance.

**Preventive Maintenance:**

The fourth maintenance activity occurs when software is changed to improve future maintainability or reliability, or to provide a better basis for future enhancements. Often called preventive maintenance, this activity is characterized by reverse engineering and re-engineering techniques.

**MODULES:**

* Image Aquistation
* Preprocessing
* Feature Extraction
* Classification
* Performance Analysis

**MODULE DESCRIPTION**

**Image Aquistation**

* Input image is the image which contains the RGB color channels. The input image is known as the dataset image.
* The input image is always taken from the dataset image, so that the intensity and the pixel value of all the images are nearly equal to one another.
* The input image which is from the dataset will give the better result than the other images which is taken from the other web pages.
* The accuracy of the process is increased due to the selection of the dataset image.

**Preprocessing:**

* **Image Resize**
* In computer graphics and digital imaging, scaling refers to the resizing of a digital image. In video technology, the magnification of digital material is known as up-scaling or resolution enhancement.
* When scaling a vector graphic image, the graphic primitives which make up the image can be scaled using geometric transformations, without any loss of image quality. When scaling a raster graphics image, a new image with a higher or lower number of pixels must be generated.

**Feature Extraction:**

* + In machine learning, pattern recognition and in image processing, feature extraction starts from an initial set of measured data and builds derived values (features) intended to be informative and non-redundant, facilitating the subsequent learning and generalization steps, and in some cases leading to better human interpretations.
  + Feature extraction is related to dimensionality reduction.
  + Feature extraction in image processing refers to the process of identifying and extracting meaningful information or patterns from raw image data.
  + It plays a crucial role in various computer vision tasks, such as object detection, image classification, and facial recognition. Feature extraction involves transforming high-dimensional image data into a lower-dimensional representation, making it more manageable and suitable for analysis. Common techniques for feature extraction include edge detection, corner detection, texture analysis, and methods based on convolutional neural networks (CNNs) and RNN. These extracted features serve as the foundation for subsequent tasks, enabling the development of robust and efficient image-based applications.
  + One of the key challenges in feature extraction is selecting the right set of features that capture the most relevant information while reducing redundancy. This process often requires domain-specific knowledge and experimentation. Feature extraction is an essential step in the image processing pipeline, as it simplifies the complexity of images and facilitates the development of machine learning models that can make sense of visual data, enabling a wide range of applications across industries, from medical imaging to autonomous vehicles.

**RNN Algorithm:**

Recurrent Neural Networks (RNNs) are a class of artificial neural networks designed to handle sequential data by maintaining a hidden state that captures temporal dependencies. The key innovation in RNNs is the ability to retain information from previous time steps, allowing them to model sequences, such as time series, natural language, and more. In an RNN, the hidden state at each time step is updated based on both the current input and the previous hidden state, making it well-suited for tasks where context and order matter. However, traditional RNNs have limitations, such as the vanishing gradient problem, which can make them less effective for capturing long-range dependencies in sequences.

To address the limitations of traditional RNNs, variants like Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) were introduced. LSTMs, for example, include specialized gates that control the flow of information, making them capable of capturing long-range dependencies and mitigating the vanishing gradient problem. RNNs find application in a wide range of tasks, including natural language processing, speech recognition, and time series prediction. Their ability to model sequential data makes them a fundamental tool in the field of deep learning.

Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks become part of a Recurrent Neural Network (RNN) when they are integrated into an architecture that allows them to operate on sequential data. In this context, CNNs are typically used for spatial feature extraction from each time step in the sequence, while LSTMs capture temporal dependencies by processing the sequences of features generated by the CNN. By combining these two neural network architectures, the resulting hybrid model, known as a Convolutional LSTM (ConvLSTM), becomes an RNN capable of effectively handling both spatial and temporal aspects of sequential data, making it well-suited for tasks such as video analysis, action recognition, and spatiotemporal data modeling.

**Classification:**

* Multi Classifier System
* In machine learning, multiclass or multinomial classification is the problem of classifying instances into one of the more than two classes (classifying instances into one of the two classes is called binary classification).
* While some classification algorithms naturally permit the use of more than two classes, others are by nature binary algorithms; these can, however, be turned into multinomial classifiers by a variety of strategies.
* Multiclass classification should not be confused with multi-label classification, where multiple labels are to be predicted for each instance.

**Performance Analysis**

* Precision and Recall:
  + In information retrieval, precision is a measure of result relevancy, while recall is a measure of how many truly relevant results are returned.
  + A high area under the curve represents both high recall and high precision, where high precision relates to a low false positive rate, and high recall relates to a low false negative rate.
  + High scores for both show that the classifier is returning accurate results (high precision), as well as returning a majority of all positive results (high recall).
  + A system with high recall but low precision returns many results, but most of its predicted labels are incorrect when compared to the training labels.
  + A system with high precision but low recall is just the opposite, returning very few results, but most of its predicted labels are correct when compared to the training labels.
  + An ideal system with high precision and high recall will return many results, with all results labeled correctly.
  + Equation for Precision:



* + Equation for Recall:



* F Measure:
  + A measure that combines precision and recall is the harmonic mean of precision and recall, the traditional F-measure or balanced F-score:
  + This measure is approximately the average of the two when they are close, and is more generally the harmonic mean which for the case of two numbers coincides with the square of the geometric mean divided by the arithmetic mean.
  + There are several reasons that the F-score can be criticized in particular circumstances due to its bias as an evaluation metric.
* Equation:



**SYSTEM REQUIREMENTS**

**SOFTWARE REQUIREMENTS:**

OS : Windows

Software : Anaconda 3 (SPYDER 3.7)

**HARDWARE REQUIREMENTS:**

Processor: Intel Pentium.

RAM: 8GB

**SOFTWARE DESCRIPTION:**

**Python**

Python is one of those rare languages which can claim to be both *simple* and powerful. You will find yourself pleasantly surprised to see how easy it is to concentrate on the solution to the problem rather than the syntax and structure of the language you are programming in. The official introduction to Python is Python is an easy to learn, powerful programming language. It has efficient high-level data structures and a simple but effective approach to object-oriented programming. Python's elegant syntax and dynamic typing, together with its interpreted nature, make it an ideal language for scripting and rapid application development in many areas on most platforms. I will discuss most of these features in more detail in the next section.

## **Features of Python**

### Simple

Python is a simple and minimalistic language. Reading a good Python program feels almost like reading English, although very strict English! This pseudo-code nature of Python is one of its greatest strengths. It allows you to concentrate on the solution to the problem rather than the language itself.

### Easy to Learn

As you will see, Python is extremely easy to get started with. Python has an extraordinarily simple syntax, as already mentioned.

### Free and Open Source

Python is an example of a FLOSS (Free/Libré and Open Source Software). In simple terms, you can freely distribute copies of this software, read its source code, make changes to it, and use pieces of it in new free programs. FLOSS is based on the concept of a community which shares knowledge. This is one of the reasons why Python is so good - it has been created and is constantly improved by a community who just want to see a better Python.

### High-level Language

When you write programs in Python, you never need to bother about the low-level details such as managing the memory used by your program, etc.

### Portable

Due to its open-source nature, Python has been ported to (i.e. changed to make it work on) many platforms. All your Python programs can work on any of these platforms without requiring any changes at all if you are careful enough to avoid any system-dependent features.

You can use Python on GNU/Linux, Windows, FreeBSD, Macintosh, Solaris, OS/2, Amiga, AROS, AS/400, BeOS, OS/390, z/OS, Palm OS, QNX, VMS, Psion, Acorn RISC OS, VxWorks, PlayStation, Sharp Zaurus, Windows CE and PocketPC!

You can even use a platform like [Kivy](http://kivy.org) to create games for your computer and for iPhone, iPad, and Android.

### Interpreted

This requires a bit of explanation.

A program written in a compiled language like C or C++ is converted from the source language i.e. C or C++ into a language that is spoken by your computer (binary code i.e. 0s and 1s) using a compiler with various flags and options. When you run the program, the linker/loader software copies the program from hard disk to memory and starts running it.

Python, on the other hand, does not need compilation to binary. You just run the program directly from the source code. Internally, Python converts the source code into an intermediate form called bytecodes and then translates this into the native language of your computer and then runs it. All this, actually, makes using Python much easier since you don't have to worry about compiling the program, making sure that the proper libraries are linked and loaded, etc. This also makes your Python programs much more portable, since you can just copy your Python program onto another computer and it just works!

### Object Oriented

Python supports procedure-oriented programming as well as object-oriented programming. In procedure-oriented languages, the program is built around procedures or functions which are nothing but reusable pieces of programs. In object-oriented languages, the program is built around objects which combine data and functionality. Python has a very powerful but simplistic way of doing OOP, especially when compared to big languages like C++ or Java.

### Extensible

If you need a critical piece of code to run very fast or want to have some piece of algorithm not to be open, you can code that part of your program in C or C++ and then use it from your Python program.

### Embeddable

You can embed Python within your C/C++ programs to give scripting capabilities for your program's users.

### Extensive Libraries

The Python Standard Library is huge indeed. It can help you do various things involving regular expressions, documentation generation, unit testing, threading, databases, web browsers, CGI, FTP, email, XML, XML-RPC, HTML, WAV files, cryptography, GUI (graphical user interfaces), and other system-dependent stuff. Remember, all this is always available wherever Python is installed. This is called the Batteries Included philosophy of Python.

Besides the standard library, there are various other high-quality libraries which you can find at the Python Package Index.

**TESTING PRODUCTS:**

System testing is the stage of implementation, which aimed at ensuring that system works accurately and efficiently before the live operation commence. Testing is the process of executing a program with the intent of finding an error. A good test case is one that has a high probability of finding an error. A successful test is one that answers a yet undiscovered error.

Testing is vital to the success of the system. System testing makes a logical assumption that if all parts of the system are correct, the goal will be successfully achieved. . A series of tests are performed before the system is ready for the user acceptance testing. Any engineered product can be tested in one of the following ways. Knowing the specified function that a product has been designed to from, test can be conducted to demonstrate each function is fully operational. Knowing the internal working of a product, tests can be conducted to ensure that “al gears mesh”, that is the internal operation of the product performs according to the specification and all internal components have been adequately exercised.

**PYTHON and images**

* The help in PYTHON is very good, use it!
* An image in PYTHON is treated as a matrix
* Every pixel is a matrix element
* All the operators in PYTHON defined on Matrices can be used on images: +, -, \*, /, ^, sqrt, sin, cos etc.
* **PYTHON can import/export several image formats**
  + BMP (Microsoft Windows Bitmap)
  + GIF (Graphics Interchange Files)
  + HDF (Hierarchical Data Format)
  + JPEG (Joint Photographic Experts Group)
  + PCX (Paintbrush)
  + PNG (Portable Network Graphics)
  + TIFF (Tagged Image File Format)
  + XWD (X Window Dump)
  + PYTHON can also load raw-data or other types of image data
* **Data types in PYTHON**
  + Double (64-bit double-precision floating point)
  + Single (32-bit single-precision floating point)
  + Int32 (32-bit signed integer)
  + Int16 (16-bit signed integer)
  + Int8 (8-bit signed integer)
  + Uint32 (32-bit unsigned integer)
  + Uint16 (16-bit unsigned integer)
  + Uint8 (8-bit unsigned integer)

**Images in PYTHON**

Binary images: {0, 1}

• Intensity images: [0, 1] or uint8, double etc.

• RGB images: m-by-n-by-3

• Indexed images: m-by-3 color map

• Multidimensional images m-by-n-by-p (p is the number of layers)

**IMAGE TYPES IN PYTHON**

Outside PYTHON images may be of three types i.e. black & white, grey scale and colored. In PYTHON, however, there are four types of images. Black & White images are called binary images, containing 1 for white and 0 for black. Grey scale images are called intensity images, containing numbers in the range of 0 to 255 or 0 to 1. Colored images may be represented as RGB Image or Indexed Image.

In RGB Images there exist three indexed images. First image contains all the red portion of the image, second green and third contains the blue portion. So for a 640 x 480 sized image the matrix will be 640 x 480 x 3. An alternate method of colored image representation is Indexed Image. It actually exist of two matrices namely image matrix and map matrix. Each color in the image is given an index number and in image matrix each color is represented as an index number. Map matrix contains the database of which index number belongs to which color.

**IMAGE TYPE CONVERSION**

* RGB Image to Intensity Image (rgb2gray)
* RGB Image to Indexed Image (rgb2ind)
* RGB Image to Binary Image (im2bw)
* Indexed Image to RGB Image (ind2rgb)
* Indexed Image to Intensity Image (ind2gray)
* Indexed Image to Binary Image (im2bw)
* Intensity Image to Indexed Image (gray2ind)
* Intensity Image to Binary Image (im2bw)
* Intensity Image to RGB Image (gray2ind, ind2rgb)

**Key Features**

* High-level language for technical computing
* Development environment for managing code, files, and data
* Interactive tools for iterative exploration, design, and problem solving
* Mathematical functions for linear algebra, statistics, Fourier analysis, filtering, optimization, and numerical integration
* 2-D and 3-D graphics functions for visualizing data
* Tools for building custom graphical user interfaces

Functions for integrating PYTHON based algorithms with external applications and languages, such as C, C++, FORTRAN, Java, COM, and Microsoft Excel.

**FEASIBILITY STUDY**

The feasibility study is carried out to test whether the proposed system is worth being implemented. The proposed system will be selected if it is best enough in meeting the performance requirements.

The feasibility carried out mainly in three sections namely.

**•** Economic Feasibility

• Technical Feasibility

• Behavioral Feasibility

**Economic Feasibility**

Economic analysis is the most frequently used method for evaluating effectiveness of the proposed system. More commonly known as cost benefit analysis. This procedure determines the benefits and saving that are expected from the system of the proposed system. The hardware in system department if sufficient for system development.

**Technical Feasibility**

This study center around the system’s department hardware, software and to what extend it can support the proposed system department is having the required hardware and software there is no question of increasing the cost of implementing the proposed system. The criteria, the proposed system is technically feasible and the proposed system can be developed with the existing facility.

**Behavioral Feasibility**

People are inherently resistant to change and need sufficient amount of training, which would result in lot of expenditure for the organization. The proposed system can generate reports with day-to-day information immediately at the user’s request, instead of getting a report, which doesn’t contain much detail.

**System Implementation**

Implementation of software refers to the final installation of the package in its real environment, to the satisfaction of the intended users and the operation of the system. The people are not sure that the software is meant to make their job easier.

* The active user must be aware of the benefits of using the system
* Their confidence in the software built up
* Proper guidance is impaired to the user so that he is comfortable in using the application

Before going ahead and viewing the system, the user must know that for viewing the result, the server program should be running in the server. If the server object is not running on the server, the actual processes will not take place.

**User Training**

To achieve the objectives and benefits expected from the proposed system it is essential for the people who will be involved to be confident of their role in the new system. As system becomes more complex, the need for education and training is more and more important. Education is complementary to training. It brings life to formal training by explaining the background to the resources for them. Education involves creating the right atmosphere and motivating user staff. Education information can make training more interesting and more understandable.

**Training on the Application Software**

After providing the necessary basic training on the computer awareness, the users will have to be trained on the new application software. This will give the underlying philosophy of the use of the new system such as the screen flow, screen design, type of help on the screen, type of errors while entering the data, the corresponding validation check at each entry and the ways to correct the data entered. This training may be different across different user groups and across different levels of hierarchy.

**Operational Documentation**

Once the implementation plan is decided, it is essential that the user of the system is made familiar and comfortable with the environment. A documentation providing the whole operations of the system is being developed. Useful tips and guidance is given inside the application itself to the user. The system is developed user friendly so that the user can work the system from the tips given in the application itself.

**System Maintenance**

The maintenance phase of the software cycle is the time in which software performs useful work. After a system is successfully implemented, it should be maintained in a proper manner. System maintenance is an important aspect in the software development life cycle. The need for system maintenance is to make adaptable to the changes in the system environment. There may be social, technical and other environmental changes, which affect a system which is being implemented. Software product enhancements may involve providing new functional capabilities, improving user displays and mode of interaction, upgrading the performance characteristics of the system. So only thru proper system maintenance procedures, the system can be adapted to cope up with these changes. Software maintenance is of course, far more than “finding mistakes”.

**Corrective Maintenance**

The first maintenance activity occurs because it is unreasonable to assume that software testing will uncover all latent errors in a large software system. During the use of any large program, errors will occur and be reported to the developer. The process that includes the diagnosis and correction of one or more errors is called Corrective Maintenance.

**Adaptive Maintenance**

The second activity that contributes to a definition of maintenance occurs because of the rapid change that is encountered in every aspect of computing. Therefore Adaptive maintenance termed as an activity that modifies software to properly interfere with a changing environment is both necessary and commonplace.

**Perceptive Maintenance**

The third activity that may be applied to a definition of maintenance occurs when a software package is successful. As the software is used, recommendations for new capabilities, modifications to existing functions, and general enhancement are received from users. To satisfy requests in this category, Perceptive maintenance is performed. This activity accounts for the majority of all efforts expended on software maintenance.

**Preventive Maintenance**

The fourth maintenance activity occurs when software is changed to improve future maintainability or reliability, or to provide a better basis for future enhancements. Often called preventive maintenance, this activity is characterized by reverse engineering and re-engineering techniques

**LITERATURE SURVEY**

**1. Title: CERVIX CANCER DIAGNOSIS FROM PAP SMEAR IMAGES USING STRUCTURE BASED SEGMENTATION AND SHAPE ANALYSIS (2012)**

**Author: Lipi B. Mahanta, 2 Dilip Ch. Nath, 1 Chandan Kr. Nath**

This work presents an approach for analysis of PAP smear images of cervical region based on cell nuclei distribution and shape and size analysis. PAP smear test is an efficient and easy procedure to detect any abnormality in cervical cells. But human observation is not always satisfying and it is a tedious task to manually analyze a large number of PAP smear images. The purpose of this study is to automate the screening process and to provide specific statistical data which will be helpful for detecting abnormalities in cervical region. The proposed approach is implemented in MATLAB®, a high level, interactive environment for data visualization/analysis/computation. The MATLAB® Image Processing Toolbox was used to segment the digital images and calculate various statistical data. By comparing cell nuclei distribution and taking into account the shape and size features MATLAB® can be programmed to distinguish normal cervical cell to questionable  
ones.

The digitized images are coloured in RGB mode. Matrix corresponding to colour image is three dimensional and hence difficult to process. The images in RGB mode have three colour components and it is a tedious task to segment using those colour component. So due to the fact that gray level image are easy to process we converted the images to gray level by using the rgb2gray unction inMATLAB. Then we enhance the contrast of the images using histogram analysis using imadjust function in MATLAB.

The cell nuclei are darker than the surrounding cytoplasm and all the cell nuclei belonging to a class tend to have same gray level. We created the histogram for  
each image to view he density distribution of the different shades of gray. Based on the fact that cell nuclei are darker, we filtered out the light areas and created a  
uniform background. Then using structuring element function, we created binary images containing only the cell nuclei. We considered only round shape for segmenting cell nuclei, but all the nuclei are not round. So, the binary images are again processed with dilation and filling function. As a result of applying dilation and erosion, extra parts which were not part of the nuclei were removed and the boundary of the cell nuclei became prominent. Then we apply filling operation to create a uniform intensity level inside the cell nuclei. An effective method to identify and classify cervical cancer is becoming increasingly needed due to the fact that early detection and a decision of correct therapy may save the patient. The statistical data can be used as benchmark to flag normal or questionable sample while the pathologist looks at the slide under a microscope which will be highly time saving. Some ideas for future enhancement includes: to design an interactive system where a pathologist can feed his own grayscale threshold or to automate the process by computer using histogram or fuzzy logic.

**Advantages:**

* The nuclei and cytoplasm of the cell is well distinguished due to the increase in contrast level of the image.

**Disadvantages:**

* Though it have some enhancement can be to establish cutoff values between normal and abnormal values and to classify the abnormal values according to stage of the cancer.

**2. Title: MORPHED SOBEL APPROACH FOR DETECTING CANCER CELLS IN LUNGS (2014)**

**Author: Vasu Kalaria, 2Priyadarsini S**

One of the gravest health problem has been lung cancer in the world in the present scenario. Automatic detection of lung cancer has provided a cutting-edge in the field of medical technology. This is because identifying cancer cells in lungs has been a tedious task as scarce or probably no symptoms are obtained in the premature stage of the disease. This project was undertaken to contribute in the field by using different techniques to identify and optimize the occurrence of cancer cells in lungs. Not only more treatment options are available, but also the increase in survival rates is observed when cancer patients are treated in their premature stages. Change in gradient provides an excellent way to identify objects present in an image. Morphing techniques play a very important role in the process of finding gradient in  
the images. Any edge detection methods can be applied on any type of medically computed images, like CT scan, X-Ray, MRI or Ultrasound; in which the objects inside the lungs are obtained. Here the use of Computed Tomography scan  
images have been made because these images are exceptional in identifying nodules as small as 2-3 mm in lungs. This technique was developed using MATLAB 8.1. A new approach named Morphed Sobel Approach has been proposed which  
makes use of Sobel edge detection algorithm along with a handful of Morphological operators to reduce the noise level as much as possible, and provide better image output that can be compared with the original image and the difference can be  
viewed with the naked eye. The framework will make use of any color type of CT scan image, i.e., RGB, Gray-scale or Binary, that contains some visualized objects by naked eye. Application of Morphological operators used day-to-day in  
every digital image processing operations. Moreover the paper provides the implementation of modified edge detection process which consists of Sobel followed after Morphing and comparison of the results obtained by the proposed method with the already existing method.

It is clearly shown that if the noise level is reduced considerably before enhancing the image, the results obtained are very convincing. The morphed sobel approach is useful in eliminating the noise level because it first uses Gaussian filter, then applies soft threshold for smoothening the image and at last makes use of binarization. It is not possible to detect edge completely by using a single fixed scale edge detection operator. Classical edge detection techniques are extremely sensitive to noise because neither the filtering is done nor the smoothening of the image is done. As a result of which a good edge detection technique should have a better noise immunity.

The new approach of edge detection named Morphed Sobel Approach was proposed in this paper. The proposed technique took into account the usage of Morphological operators, Gaussian filters, de-noising as well as simple Sobel operator. The combination of which was able to produce better outputs and results as compared to the existing techniques of edge detection which suffer badly from noise. The edges that were detected by the proposed system were in good shape unlike those poorly detected edges in classical techniques of edge detection.

**Advantages:**

* The segmentation process gives the better result, and also the simulation process is also high.

**Disadvantages:**

* The threshold value given to this segmentation process will remains same, so that the accurate result value will not occur easily.

**3. Title: ADAPTIVE MULTI THRES-HOLDING FOR BREAST CANCER STEM CELL DETECTION (2013)**

**Author: Sabina1, Mrs. Nidhi**

Image analysis of cancer cells is important for breast cancer diagnosis and therapy, because it is recognized as most effective and efficient way to observe its generation. In this paper we are going to present a noble method to detect breast cancer stem cell for an image, as the growth in automatic detection of breast cancer using image processing grows it attract many researchers to research and optimized the detection for breast cancer using various algorithms. I am going to use multi thresholding concept to detect breast cancer stem in biomedical images and to implement the concept of over segmentation so that no cell left behind.

Various methods have been proposed for automating breast cancer detection. Many morphological feature based methods are extracted based on the segmentation of biopsy images such as by counting the components after binarzation and by measuring dimensions of segmented region after Hough transform. Such morphological results are doubtful and solely depend upon quality of segmentation, which is mostly degraded by image noise. Thus, I propose a method based on pixel-wise posterior probabilities of co-occurrence of cell components to improve  
the accuracy of breast cancer detection from stained tissues of biopsy images. Although the previous approaches have shown some advantages, but there are still some shortcomings. Firstly, Level set and Active contour takes more processing time and thus time-consuming. In both approaches, synchronously segmentation inaccuracy cannot be avoided and overcame. Secondly, double threshold provides simple and easy way to handle segmentation problem instead of lengthy equations and complicated mathematic theory. But still we can see some drawbacks in this approach too in aspect of segmentation. So, I here propose a new approach Adaptive Multi-Thresholding detection of breast cancer stem cell which not only  
aim to the introduction and analysis of adaptive multi-thresholding method but also for the purpose of comparison of different ways.

Here, a finite-difference time-domain (FDTD) simulations of acoustic models and electromagnetic (EM) are targeted to investigate acoustic pressure generated by spherical breast tumor. Simulation results show quantative relationship  
between input power and output acoustic pressure. In this paper, accurate EM model of the TAI system including fluidic environment [22], feeding antenna, and 3D sample are studied and developed in a broadband frequency range.  
In recent years, many TAI demonstrations have been reported [23-24]. 2D imaging of breast tissue phantoms has been achieved. Acoustic wave propagation established in previous works has not been much reported to work on the  
electromagnetic (EM) modeling of TAI instrument and its quantative correlation with acoustic output. Thus, microwave-induced thermal acoustic imaging (TAI), combining the high contrast of microwave imaging and high resolution of ultrasound imaging, is a promising candidate of breast cancer detection. Considering all the approaches and models above adaptive multi thresholding for breast cancer stem cell detection can give good results over segmentation.

**Advantages:**

* In this project the cancer cells are detected in the previous stage, so that this method is also applicable for the initial stage also.

**Disadvantages:**

* The further work will be done by optimizing multiple random threshold values and taking image into MATLAB workspace for an efficient image recognition and highest accuracy.

**4. Title: EXTRACTION OF CANCER CELLS FROM MRI PROSTATE**

**IMAGE USING MATLAB (2012)**

**Author: N.Gopinath**

Medical Image Processing is one of the most challenging and emerging topics in today’s research field. Processing of Magnetic Resonance Imaging (MRI) is one of the parts in this field. In recent years, multispectral MRI has emerged as an alternative to Ultrasound (US) image modality for clear identification of cancer in Breast, Prostate and Liver etc,. In order to analyze a disease, Physicians consider MR imaging modality is the most efficient one for identification of cancer present in various organs. Therefore, analysis on MR imaging is required for efficient disease diagnosis. This paper describes the proposed strategy to detect and extraction of Prostate cancer cells from patient’s MRI scan image of the Prostate organ. This proposed method incorporates with some noise removal functions, segmentation and morphological functions which are considered to be the basic concepts of Image Processing. Detection and extraction of cancer cells from MRI Prostate image is done by using the MATLAB software.

Image pre-processing is the term for operations on images at the lowest level of abstraction. These operations do not increase image information content, but they decrease it if entropy is an information measure. The aim of pre-processing is an improvement of the image data that suppresses undesired distortions or enhances some image features relevant for further processing and analysis task. Pre-processing of MRI images is the primary step in image analysis which perform image enhancement and noise reduction techniques which are used to  
enhance the image quality, then some morphological operations are applied to detect the cancer cells in the image. The morphological operations are basically applied on some assumptions about the size and shape of the cancer and in the end the cancer cells are mapped onto the original gray scale image with 255 intensity to make  
visible the cancer in the image. The algorithm has been tried on number of patients MRI data of Prostate cancer images.

Image segmentation is the process of partitioning the digital/ medical image into multiple regions that can be associated with the properties of one or more criterion. It is an initial and vital step in pattern recognition - a series  
of processes aimed at overall image understanding. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain visual characteristics. The result of image segmentation is a set of segments that collectively cover the entire image, or  
a set of contours extracted from the image. Each of the pixels in a region is similar with respect to some characteristic or computed property, such as color, intensity, or texture. Adjacent regions are significantly different with respect to the same characteristic(s).Thresholding is the simplest Image Segmentation method. This method is based on threshold value to convert the gray level image into a binary image. The key of this method is to select the threshold value (or values when multiple-levels are selected).

**Advantages:**

* The Gaussian filter and morphological operation is combined to give the better enhancement and give the noise free output.

**Disadvantages:**

* In this project some of the drawbacks is, contour is not done to separate the region hence the noise reduction only undertaken.

**5. Title: AUTOMATIC DETECTION SYSTEM OF CERVICAL CANCER CELLS USING COLOR INTENSITY CLASSIFICATI-ON**

**Author: Eko supriyanto, Ur azureen m. pista,lukman hakim ismail, Bustanur rosidi, tati latifah mengko**

The conventional Pap smear has been undeniably responsible in reducing the number of incidence and mortality of cervical cancer. However, few concerns have arisen such as the shortage of skilled and experienced pathologists and the increasing workload as a result of more individuals having gained access to preventive health care which eventually will make the reviewing procedure becomes time consuming and highly prone to human errors. In order to solve this problem, an automated detection system of cervical cancer cells has been developed. The detection of cervical cancer cells is based on the morphology of the cells and level set operations. Test result shows, that by using color intensity classification the system is able to differentiate between normal and cancerous cells. This system will hopefully help the pathologist to reduce the work-load and minimize human error while maintaining and improving the accuracy of the system.

Resources used in this project include hardware/ modalities, software, and content & database. The hardware device used was Olympus BX-51 microscope.  
This microscope was used for data collection of cervical cells images. It examined the desired section of microscopic slide taken from Pap smear test. The desired  
section has been captured and stored in a computer using imaging software (Olympus). The software used to process the images was MATLAB with Image Processing Toolbox and the system’s prototype was designed using  
graphical user interface (GUI). All these data as well patient data bases were collected from National Cancer Hospital Dharmais, Jakarta, Indonesia.  
To achieve its objective this project followed these stepwise procedures: data acquisition, image processing, feature extraction, segmentation for cells’ detection and software development using MATLAB.

The objective of image analysis is to measure the elements’ properties such as color intensity level, area, and centroid using ‘regionprops’ function. These  
properties will determine whether the element is cervical cell or not. Moreover, ‘bwboundaries’ function was also used to return a cell array, where each cell contains the row and column coordinates for an element in the image. The coordinates returned by ‘bwboundaries’ were then used to plot the borders of all elements in the image. Lastly, all elements were numbered from top to bottom,  
then from left to right. The result is shown in the figure below. The distinctive differences of color intensity distributions between normal and cancerous cells have been successfully applied to be used to characterize cancerous cervical cancer cells. Test result shows, that the detection system in this study is able to differentiate between normal and cancerous cells by using color intensity classification. We are optimistic that this system will help in reducing the pathologists’ work-load as well as in minimizing possible human error occurring in the analyzing of conventional Pap smear, while maintaining and improving the accuracy of the system.

**Advantages:**

* This system will hopefully help the pathologist to reduce the work-load and minimize human error while maintaining and improving the accuracy of the system.

**Disadvantages:**

* Although there are number of steps to enhance the image, this project have some drawbacks. In this project the stages of cancer will not be able to detect.

**REFERENCES**

[1] E. W. Elston and I. O. Ellis, ‘‘Method for grading breast cancer,’’ J. Clin. Pathol., vol. 46, no. 2, pp. 189–190, 1993.

[2] M. M. Dundar et al., ‘‘Computerized classification of intraductal breast lesions using histopathological images,’’ IEEE Trans. Biomed. Eng., vol. 58, no. 7, pp. 1977–1984, Jul. 2011.

[3] C. D. Malon and E. Cosatto, ‘‘Classification of mitotic figures with convolutional neural networks and seeded blob features,’’ J. Pathol. Inform., vol. 4, no. 1, p. 9, 2013.

[4] H. Chen, Q. Dou, X. Wang, J. Qin, and P. A. Heng, ‘‘Mitosis detection in breast cancer histology images via deep cascaded networks,’’ in Proc. 13th AAAI Conf. Artif. Intell., 2016, pp. 1160–1166.

[5] (2014). MITOS, ICPR 2014 Contest, IPAL UMI CNRS Lab Std. [Online]. Available: <http://ipal.cnrs.fr/ICPR2014>

[6] L. Roux et al., ‘‘Mitosis detection in breast cancer histological images An ICPR 2012 contest,’’ J. Pathol. Inform., vol. 4, no. 1, p. 8, 2013.

[7] P. W. Barlow, ‘‘Changes in chromatin structure during the mitotic cycle,’’ Protoplasma, vol. 91, no. 2, pp. 207–211, Jun. 1977.

[8] S. Lankton and A. Tannenbaum, ‘‘Localizing region-based active contours,’’ IEEE Trans. Image Process., vol. 17, no. 11, pp. 2029–2039, Nov. 2008.

[9] A. H. Gandomi and A. H. Alavi, ‘‘Krill herd: A new bio-inspired optimization algorithm,’’ Commun. Nonlinear Sci. Numer. Simul., vol. 17, no. 12, pp. 4831–4845, Dec. 2012.

[10] B. F. Buxton, W. B. Langdon, and B. J. Barrett, ‘‘Data fusion by intelligent classifier combination,’’ Meas. Control-London Inst. Meas. Control, vol. 34, no. 8, pp. 229–234, Oct. 2001.

[11] J. A. M. Beliën, J. P. A. Baak, P. J. van Diest, and A. H. M. Van Ginkel, ‘‘Counting mitoses by image processing in Feulgen stained breast cancer sections: The influence of resolution,’’ Cytometry A, vol. 28, no. 2, pp. 135–140, Jun. 1997.

[12] A. M. Khan, H. El-Daly, and N. M. Rajpoot, ‘‘A Gamma-Gaussian mixture model for detection of mitotic cells in breast cancer histopathology images,’’ in Proc. 21st Int. Conf. Pattern Recognit. (ICPR), Nov. 2012, pp. 149–152.

[13] F. B. Tek et al., ‘‘Mitosis detection using generic features and an ensemble of cascade adaboosts,’’ J. Pathol. Inform., vol. 4, no. 1, p. 12, 2013.

[14] V. Roullier, O. Lézoray, V.-T. Ta, and A. Elmoataz, ‘‘Multi-resolution graph-based analysis of histopathological whole slide images: Application to mitotic cell extraction and visualization,’’ Comput. Med. Imag. Graph., vol. 35, nos. 7–8, pp. 603–615, Oct./Dec. 2011.

[15] C. Sommer, L. Fiaschi, F. A. Hamprecht, and D. W. Gerlich, ‘‘Learningbased mitotic cell detection in histopathological images,’’ in Proc. 21st Int. Conf. Pattern Recognit. (ICPR), Nov. 2012, pp. 2306–2309.

[16] O. Sertel, U. V. Catalyurek, and M. N. Gurcan, ‘‘Computer-aided prognosis of neuroblastoma: Detection of mitosis and karyorrhexis cells in digitized histological images,’’ in Proc. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC), Sep. 2009, pp. 1433–1436.

[17] K. S. Beevi, M. S. Nair, and G. R. Bindu, ‘‘Detection of mitotic nuclei in breast histopathology images using localized ACM and random kitchen sink based classifier,’’ in Proc. IEEE 38th Annu. Int. Conf. Eng. Med. Biol. Soc. (EMBC), Aug. 2016, pp. 2435–2439.

[18] H. Irshad et al., ‘‘Automated mitosis detection using texture, SIFT features and HMAX biologically inspired approach,’’ J. Pathol. Inform., vol. 4, no. 2, p. 12, 2013.

[19] S. Doyle, S. Agner, A. Madabhushi, M. Feldman, and J. Tomaszewski, ‘‘Automated grading of breast cancer histopathology using spectral clus tering with textural and architectural image features,’’ in Proc. 5th IEEE Int. Symp. Biomed. Imag., Nano Macro (ISBI), May 2008, pp. 496–499.

[20] K. S. Beevi and G. R. Bindu, ‘‘Analysis of nuclei detection with stain normalization in histopathology images,’’ Indian J. Sci. Technol., vol. 8, no. 23, Sep. 2015, pp. 547–551.

[21] M. Veta, P. J. van Diest, and J. P. W. Pluim, ‘‘Detecting mitotic fig ures in breast cancer histopathology images,’’ Proc. SPIE, vol. 8676, pp. 867607-1–867607-7, Mar. 2013.

[22] V. Anari, P. Mahzouni, and R. Amirfattahi, ‘‘Computer-aided detection of proliferative cells and mitosis index in immunohistichemically images of meningioma,’’ in Proc. 6th Iranian Conf. Mach. Vis. Image Process., Oct. 2010, pp. 1–5.

[23] M. N. Gurcan, T. Pan, H. Shimada, and J. Saltz, ‘‘Image analysis for neuroblastoma classification: Segmentation of cell nuclei,’’ in Proc. 28th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBS), Aug./Sep. 2006, pp. 4844–4847.

[24] S. Beevi, M. S. Nair, and G. R. Bindu, ‘‘Automatic segmentation and classification of mitotic cell nuclei in histopathology images based on active contour model,’’ in Proc. Int. Conf. Contemp. Comput. Inform. (IC3I), Nov. 2014, pp. 740–744.

[25] A. Paul and D. P. Mukherjee, ‘‘Mitosis detection for invasive breast cancer grading in histopathological images,’’ IEEE Trans. Image Process., vol. 24, no. 11, pp. 4041–4054, Nov. 2015.

[26] C. Lu and M. Mandal, ‘‘Toward automatic mitotic cell detection and segmentation in multispectral histopathological images,’’ IEEE J. Biomed. Health Inform., vol. 18, no. 2, pp. 594–605, Mar. 2014.

[27] H. Irshad, A. Gouaillard, L. Roux, and D. Racoceanu, ‘‘Multispectral band selection and spatial characterization: Application to mitosis detection in breast cancer histopathology,’’ Comput. Med. Imag. Graph., vol. 38, no. 5, pp. 390–402, Jul. 2014.

[28] H. Wang et al., ‘‘Cascaded ensemble of convolutional neural networks and handcrafted features for mitosis detection,’’ J. Med. Imag., vol. 1, no. 3, p. 90410B, Mar. 2014.

[29] D. C. Cireşan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, ‘‘Mitosis detection in breast cancer histology images with deep neural networks,’’ in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent., 2013, pp. 411–418.

[30] L. I. Kuncheva, J. C. Bezdek, and R. P. W. Duin, ‘‘Decision templates for multiple classifier fusion: An experimental comparison,’’ Pattern Recognit., vol. 34, no. 2, pp. 299–314, Feb. 2001.

[31] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, ‘‘On combining classifiers,’’ IEEE Trans. Pattern Anal. Mach. Intell., vol. 20, no. 3, pp. 226–239, Mar. 1998.

[32] G. Hinton, ‘‘A practical guide to training restricted Boltzmann machines,’’ Momentum, vol. 9, no. 1, p. 926, 2010.

[33] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, ‘‘Greedy layer-wise training of deep networks,’’ in Advances in Neural Information Processing Systems, vol. 19. Cambridge, MA, USA: MIT Press, 2007, p. 153.

[34] A. M. Khan, N. Rajpoot, D. Treanor, and D. Magee, ‘‘A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution,’’ IEEE Trans. Biomed. Eng., vol. 61, no. 6, pp. 1729–1738, Jun. 2014.

[35] S. Haykin and B. Widrow, Eds., Least-Mean-Square Adaptive Filters, vol. 31. Hoboken, NJ, USA: Wiley, 2003.