COVID 19 DETECTION IN XRAY IMAGES USING MACHINE LEARNING ALGORITHMS SUPPORT VECTOR MACHINES

ABSTRACT:

Covid-19 may be a speedily spreading infective agent unwellness that infects not solely humans; however, animals also are infected due to this unwellness. The standard of living of kith and kin, their health, and also the economy of a rustic square measure affected because of this deadly infective agent unwellness. Covid-19 is a common spreading unwellness, and until currently, not one country will prepare a immunizing agent for COVID-19. A clinical study of COVID-19 infected patients has shown that these varieties of patient’s square measure largely infected from a respiratory organ infection when coming back in contact with this unwellness. Chest x-ray (i.e., radiography) and chest CT square measure a more practical imaging technique for identification lunge connected issues. Still, a considerable chest x-ray may be a lower price method compared to chest CT. Deep learning is that the most successful technique of machine learning, that provides helpful analysis to review an outsized quantity of chest x-ray pictures that may critically impact on screening of Covid-19. during this work, we've taken the PA read of chest x-ray scans for covid-19 affected patients additionally as healthy patients. when clean-up up the pictures and applying information augmentation, we have used four machine learning algorithms logistic regression, naïve ayes, support vector machine, random forest examined their accuracy. to investigate the model performance, 6432 chest x-ray scans samples are collected from the Kaggle repository, out of that 5467 were used for coaching and 965 for validation. In result analysis, the Exception model provides the highest accuracy (i.e., 97.97%) for police investigation Chest X-rays pictures as compared to alternative models. This work solely focuses on possible strategies of classifying covid-19 infected patients and doesn't claim any medical accuracy.

Keywords: covid19, x ray images, machine learning, accuracy.

1.INTRODUCTION:

Covid-19 could be a severe illness issue wherever an oversized range of people lose their lives each day. This illness affects not solely a single country, and even the total world suffered as a result of of this virus illness. within the past decade, many styles of viruses (like severe acute respiratory syndromecame into the picture, however they symbolize solely many days or few months . Many scientists square measure performing on these styles of viruses, and few of them square measure diagnosed thanks to the supply of vaccines prepared by them (i.e., Scientists or researchers). within the gift time, the total world is littered with Covid-19 illness, and the most vital issue is not any single country scientists will prepare a immunizing agent for identical. Meanwhile, more predictions came into an image such as plasma medical care, X-ray pictures , and plenty of a lot of, but the precise answer of this deathly illness isn't found. Every day, folks lose their life thanks to covid-19, and the diagnostic price of this illness is incredibly high within the context of a country, state, and patients. In March 2020, X-ray pictures of healthy folks and Covid-19 infected peoples were available on-line in numerous repositories like Github, Kaggle for analysis.threatens humans at a world level and was a pestilence. To diagnose covid-19 infected patients with healthy patients is a crucial task. The chemical analysis of Covid-19 infected patients desires more precaution and should be cured below terribly strict procedures to scale back the chance of patients unaffected with covid-19. The novel coronavirus illness came initial as a throat infection, and suddenly folks sweet-faced problem in respiration. The covid-19 ill health could be a hidden enemy wherever nobody is capable of fighting. Infected patients of Covid-19square measure needed to be in isolation, do correct screening, and take adequate protection with interference to guard healthy folks. This infection is following a sequence methodthat transfers from one person to another when coming back connected with covid-19 infected persons. Hospital employees, nurses, doctors, and clinical facilities play an essential role within the designation of this epidemic. Many more strategies are applied to scale back the impact of Covid-19. Medical imaging [ is additionally a technique of analyzing and predicting the consequences of covid-19 on the form. In this, healthy folks and Covid-19 infected patients may be analyzed in parallel with the assistance of CT (Computerised Tomography) imagesand chest X-ray pictures. For tributary to Associate in Nursing analysis of Covid-19, we tend to collected uploaded information of X-ray images of healthy and covid-19 infected patients from different sources and applied 3 totally different models (InceptionV3, Xception, and ResNeXt). The analysis of this collected information is done with the assistance of ML, a machine learning tool. This work principally focuses on the employment of ML models for classifying chest X-ray pictures for coronavirus infected patients. We have tried to draw a parallel to the previous add the field and appearance for potential models of the task, which might be assessed more to prove their quality in sensible scenarios.

2.RELATED WORK:

the authors projected a framework model supported Capsule Networks to diagnose Covid-19 (i.e., COVIDCAAPS) illness with the assistance of X-ray pictures. during this proposed work, many convolution layers and capsules area unit used to overcome the matter of class-imbalance. In experimental analysis, they showed the satisfying performance of COVIDCAPS on a smaller range of trainable parameters. Authors mentioned concerning the thought-about trained model that is publicly out there on Githubfor open access. As a result, they concluded that the projected model shows accuracy ninety five.7%, whereas sensitivity is shown as ninetieth and specificity as 95.80% whereas applying a smaller range of trainable parameters. In , the authors thought-about the primary 3 cases of Covid-19 infected cases in France. Out of those 3 persons, two were diagnosed in Paris and one in Bordeaux. Before coming in touch with Covid-19 diseases, they were staying in Wuhan, China. In , the author projected a hybrid system supported artificial intelligence, that specially used machine learning and deep learning algorithms (i.e., Convolutional Neural Network (ML) mistreatment softmax classifier). The projected system is specially enforced for detective work Covid-19 cases using chest X-ray pictures. In , the authors have given a radiologic analysis of MERS (Middle East metabolism Syndrome) on novel coronavirus. They thought-about the case of a thirty year previous male patient United Nations agency suffered from diarrhoea, fever, and abdominal pain. The authors gave associate analysis of the treatment of infected persons with chest X-rays. Further, they applied this model on a collected dataset of chest X-ray and CT pictures and received improved results. Also, in , they mentioned what form of protocols area unit required to follow by hospital employees to reduce the danger of healthy patients and what precaution is required in taking care of covid-19 infected patients. In , the authors mentioned the natural event of etiology in metropolis, China. They additionally raised a question concerning the precise reason behind this epidemic. during this study, they value the traveling (via business or air) impact on covid-19.the authors applied the SVM technique to spot pneumothorax. They used an area Binary Pattern (LBP) to mine the characteristics of respiratory organ pictures. within the projected detection model, the authors used multi-scale texture segmentation by removing impurities of chest pictures for segmenting the regions of abnormal lungs. Further, this transformation was applied for a amendment of texture for locating multiple overlapping blocks. Finally, the authors used rid of boundary (with Sobel Edge detection) for locating an entire region of illness with the abnormal half. , the authors thought-about twenty one covid-19 patients’ chest CT scans in metropolis, China. The authors majorly specialize in the demonstration of covid-19 illness on human’s lungs and their impacts. Next, the authors proposed a COVID-RENet model to extract the options (i.e., edge and region-based) with ML for classification. In this work, authors acquire options by applying ML, and afterward, they used SVM to boost the performance of classification. They used 5-fold cross-validation on a collected dataset of Covid-19. This projected approach is especially applicable to a medical specialist for the first identification of Covid-19 infected patients. the authors applied a deep learning model on the collected image dataset of chest CT to spot the impacts of Covid-19 from persons noninheritable respiratory disorder and respiratory organ disease.

the authors mentioned the chest radiography (CXR) for identification of respiratory organ abnormality. They show that the medical community can have confidence CXR as a result of its full availability and reduced infection management. They used 123 front views of X-ray for the detection of Covid-19 diseases. Further, the authors mentioned the role of AI tools in healthcare. They additionally talked concerning the challenges of implementing AI tools on less dataset of X-ray pictures (which is available publically). The authors thought-about a dataset of Xrays and CT pictures from many resources and applied deep learning and transfer learning algorithms to notice Covid-19 diseases. Picklepre-trained and changed ML model has been used on a collected dataset. As a result, they showed that the pre-trained model gave ninety eight accuracy, and adjusted ML shows 94.1% accuracy of the model.

**DIGITAL IMAGE PROCESSING**

The identification of objects in an image and this process would probably start with image processing techniques such as noise removal, followed by (low-level) feature extraction to locate lines, regions and possibly areas with certain textures.

The clever bit is to interpret collections of these shapes as single objects, e.g. cars on a road, boxes on a conveyor belt or cancerous cells on a microscope slide. One reason this is an AI problem is that an object can appear very different when viewed from different angles or under different lighting. Another problem is deciding what features belong to what object and which are background or shadows etc. The human visual system performs these tasks mostly unconsciously but a computer requires skilful programming and lots of processing power to approach human performance. Manipulation of data in the form of an image through several possible techniques. An image is usually interpreted as a two-dimensional array of brightness values, and is most familiarly represented by such patterns as those of a photographic print, slide, television screen, or movie screen. An image can be processed optically or digitally with a computer.

1. **Basics of Image Processing:-**

**FUNDAMENTALS OF DIGITAL IMAGE**

**1.1 IMAGE:**

An image is a two-dimensional picture, which has a similar appearance to some subject usually a physical object or a person.

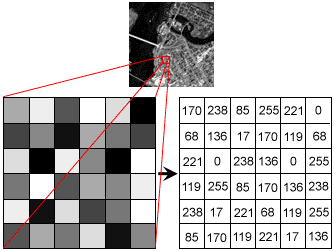
Image is a two-dimensional, such as a photograph, screen display, and as well as a three-dimensional, such as a statue. They may be captured by optical devices—such as cameras, mirrors, lenses, telescopes, microscopes, etc. and natural objects and phenomena, such as the human eye or water surfaces.

The word image is also used in the broader sense of any two-dimensional figure such as a map, a graph, a pie chart, or an abstract painting. In this wider sense, images can also be rendered manually, such as by drawing, painting, carving, rendered automatically by printing or computer graphics technology, or developed by a combination of methods, especially in a pseudo-photograph.

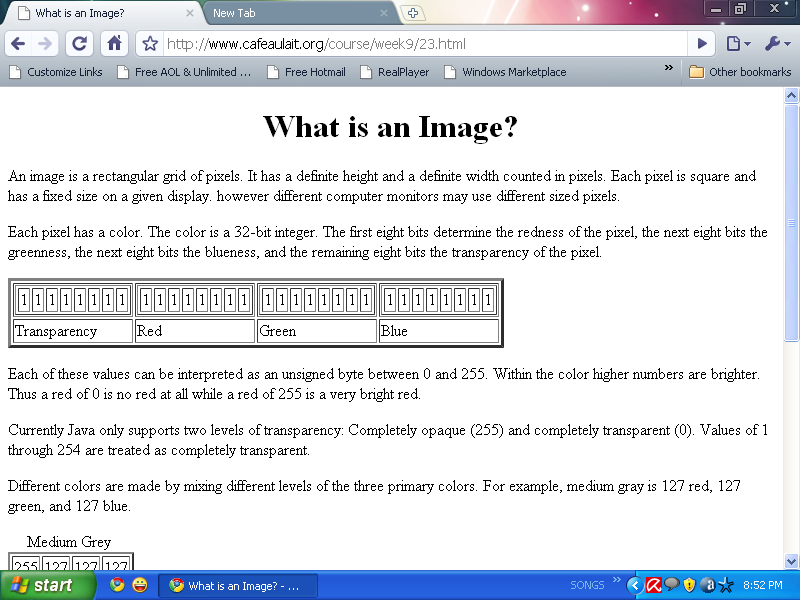
**Fig: Colour image to Gray scale Conversion Process**

An image is a rectangular grid of pixels. It has a definite height and a definite width counted in pixels. Each pixel is square and has a fixed size on a given display. However different computer monitors may use different sized pixels. The pixels that constitute an image are ordered as a grid (columns and rows); each pixel consists of numbers representing magnitudes of brightness and color.



**Fig: Gray Scale Image Pixel Value Analysis**

Each pixel has a color. The color is a 32-bit integer. The first eight bits determine the redness of the pixel, the next eight bits the greenness, the next eight bits the blueness, and the remaining eight bits the transparency of the pixel.



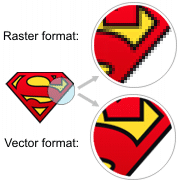
**Fig: BIT Transferred for Red, Green and Blue plane (24bit=8bit red;8-bit green;8bit blue)**

**IMAGE FILE SIZES:**

Image file size is expressed as the number of bytes that increases with the number of pixels composing an image, and the color depth of the pixels. The greater the number of rows and columns, the greater the image resolution, and the larger the file. Also, each pixel of an image increases in size when its color depth increases, an 8-bit pixel (1 byte) stores 256 colors, a 24-bit pixel (3 bytes) stores 16 million colors, the latter known as true color.Image compression uses algorithms to decrease the size of a file. High resolution cameras produce large image files, ranging from hundreds of kilobytes to megabytes, per the camera's resolution and the image-storage format capacity. High resolution digital cameras record 12 megapixel (1MP = 1,000,000 pixels / 1 million) images, or more, in true color. For example, an image recorded by a 12 MP camera; since each pixel uses 3 bytes to record true color, the uncompressed image would occupy 36,000,000 bytes of memory, a great amount of digital storage for one image, given that cameras must record and store many images to be practical. Faced with large file sizes, both within the camera and a storage disc, image file formats were developed to store such large images.

**IMAGE FILE FORMATS:**

Image file formats are standardized means of organizing and storing images. This entry is about digital image formats used to store photographic and other images. Image files are composed of either pixel or vector (geometric) data that are rasterized to pixels when displayed (with few exceptions) in a vector graphic display. Including proprietary types, there are hundreds of image file types. The PNG, JPEG, and GIF formats are most often used to display images on the Internet.



**Fig: Horizontal and Vertical Process**

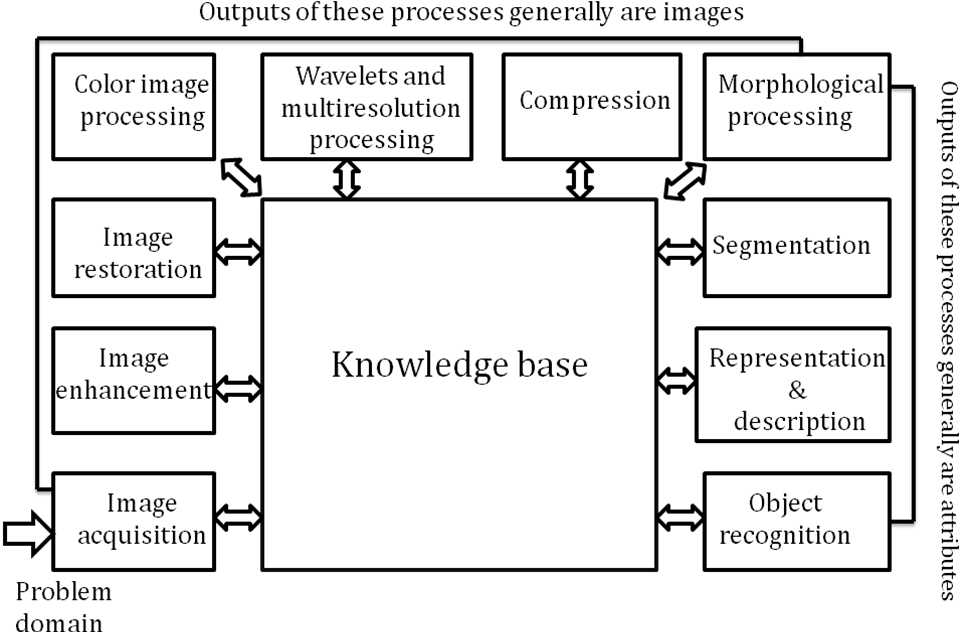
In addition to straight image formats, Metafile formats are portable formats which can include both raster and vector information. The metafile format is an intermediate format. Most Windows applications open metafiles and then save them in their own native format.

**IMAGE PROCESSING:**

Digital image processing, the manipulation of images by computer, is relatively recent development in terms of man’s ancient fascination with visual stimuli. In its short history, it has been applied to practically every type of images with varying degree of success. The inherent subjective appeal of pictorial displays attracts perhaps a disproportionate amount of attention from the scientists and also from the layman. Digital image processing like other glamour fields, suffers from myths, mis-connect ions, mis-understandings and mis-information. It is vast umbrella under which fall diverse aspect of optics, electronics, mathematics, photography graphics and computer technology. It is truly multidisciplinary endeavor ploughed with imprecise jargon.

Several factor combine to indicate a lively future for digital image processing. A major factor is the declining cost of computer equipment. Several new technological trends promise to further promote digital image processing. These include parallel processing mode practical by low cost microprocessors, and the use of charge coupled devices (CCDs) for digitizing, storage during processing and display and large low cost of image storage arrays.

**FUNDAMENTAL STEPS IN DIGITAL IMAGE PROCESSING:**

**Fig: Basics steps of image Processing**

**Image Acquisition:**

**Image Acquisition** is to acquire a digital image. To do so requires an image sensor and the capability to digitize the signal produced by the sensor. The sensor could be monochrome or color TV camera that produces an entire image of the problem domain every 1/30 sec. the image sensor could also be line scan camera that produces a single image line at a time. In this case, the objects motion past the line.



**Fig: Digital camera**

Scanner produces a two-dimensional image. If the output of the camera or other imaging sensor is not in digital form, an analog to digital converter digitizes it. The nature of the sensor and the image it produces are determined by the application.



**Fig: Mobile based Camera**

**Image Enhancement:**

**Image enhancement**is among the simplest and most appealing areas of digital image processing. Basically, the idea behind enhancement techniques is to bring out detail that is obscured, or simply to highlight certain features of interesting an image. A familiar example of enhancement is when we increase the contrast of an image because “it looks better.” It is important to keep in mind that enhancement is a very subjective area of image processing.

  
**Fig: Image enhancement process for Gray Scale Image and Colour Image using Histogram Bits**

**1.5.3 Image restoration:**

**Image restoration**is an area that also deals with improving the appearance of an image. However, unlike enhancement, which is subjective, image restoration is objective, in the sense that restoration techniques tend to be based on mathematical or probabilistic models of image degradation.



Fig: Noise image🡪 Image Enhancement

Enhancement, on the other hand, is based on human subjective preferences regarding what constitutes a “good” enhancement result. For example, contrast stretching is considered an enhancement technique because it is based primarily on the pleasing aspects it might present to the viewer, where as removal of image blur by applying a deblurring function is considered a restoration technique.

**Color image processing:**

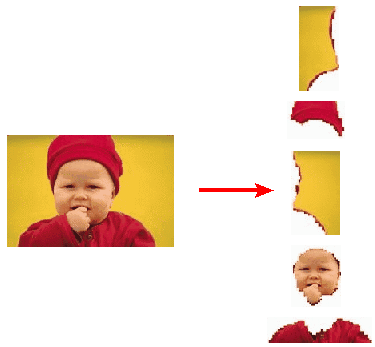
The use of color in image processing is motivated by two principal factors. First, color is a powerful descriptor that often simplifies object identification and extraction from a scene. Second, humans can discern thousands of color shades and intensities, compared to about only two dozen shades of gray. This second factor is particularly important in manual image analysis.

****

**Fig: gray Scale image 🡪 Colour Image**

**Segmentation:**

**Segmentation**procedures partition an image into its constituent parts or objects. In general, autonomous segmentation is one of the most difficult tasks in digital image processing. A rugged segmentation procedure brings the process a long way toward successful solution of imaging problems that require objects to be identified individually.



**Fig: Image Segment Process**

On the other hand, weak or erratic segmentation algorithms almost always guarantee eventual failure. In general, the more accurate the segmentation, the more likely recognition is to succeed.

Digital image is defined as a two dimensional function f(x, y), where x and y are spatial (plane) coordinates, and the amplitude of f at any pair of coordinates (x, y) is called intensity or grey level of the image at that point. The field of digital image processing refers to processing digital images by means of a digital computer. The digital image is composed of a finite number of elements, each of which has a particular location and value. The elements are referred to as picture elements, image elements, pels, and pixels. Pixel is the term most widely used.

**Image Compression**

Digital Image compression addresses the problem of reducing the amount of data required to represent a digital image. The underlying basis of the reduction process is removal of redundant data. From the mathematical viewpoint, this amounts to transforming a 2D pixel array into a statically uncorrelated data set. The data redundancy is not an abstract concept but a mathematically quantifiable entity. If n1 and n2 denote the number of information-carrying units in two data sets that represent the same information, the relative data redundancy  [2] of the first data set (the one characterized by n1) can be defined as,



Where  called as compression ratio [2]. It is defined as

= 

In image compression, three basic data redundancies can be identified and exploited: Coding redundancy, interpixel redundancy, and phychovisal redundancy. Image compression is achieved when one or more of these redundancies are reduced or eliminated. The image compression is mainly used for image transmission and storage. Image transmission applications are in broadcast television; remote sensing via satellite, air-craft, radar, or sonar; teleconferencing; computer communications; and facsimile transmission. Image storage is required most commonly for educational and business documents, medical images that arise in computer tomography (CT), magnetic resonance imaging (MRI) and digital radiology, motion pictures, satellite images, weather maps, geological surveys, and so on.

**Image Compression Model**

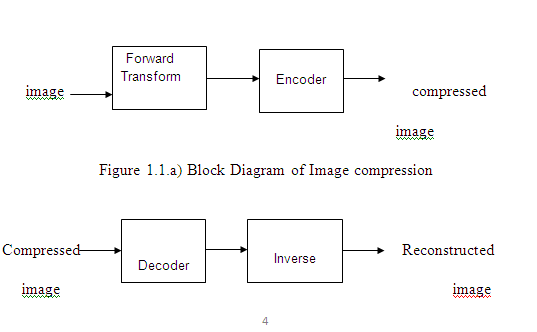


Fig:1.1b) Decompression Process for Image

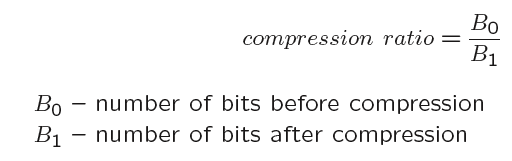
**Image Compression Types**

There are two types’ image compression techniques.

1. Lossy Image compression

2. Lossless Image compression

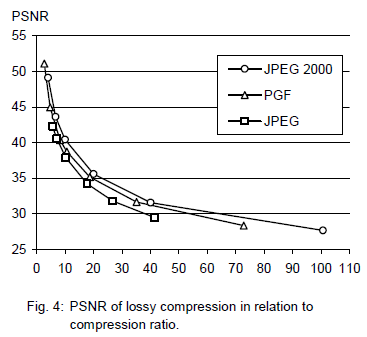
Compression ratio:



**1. Lossy Image compression :**

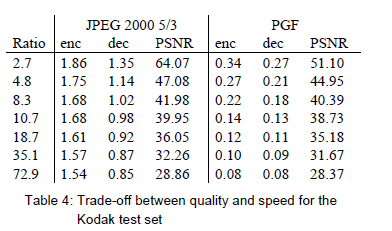
Lossy compression provides higher levels of data reduction but result in a less than perfect reproduction of the original image. It provides high compression ratio. lossy image compression is useful in applications such as broadcast television, videoconferencing, and facsimile transmission, in which a certain amount of error is an acceptable trade-off for increased compression performance. Originally, PGF has been designed to quickly and progressively decode lossy compressed aerial images. A lossy compression mode has been preferred, because in an application like a terrain explorer texture data (e.g., aerial orthophotos) is usually mid-mapped filtered and therefore lossy mapped onto the terrain surface. In addition, decoding lossy compressed images is usually faster than decoding lossless compressed images.

In the next test series we evaluate the lossy compression efficiency of PGF. One of the best competitors in this area is for sure JPEG 2000. Since JPEG 2000 has two different filters, we used the one with the better trade-off between compression efficiency and runtime. On our machine the 5/3 filter set has a better trade-off than the other. However, JPEG 2000 has in both cases a remarkable good compression efficiency for very high compression ratios but also a very poor encoding and decoding speed. The other competitor is JPEG. JPEG is one of the most popular image file formats.

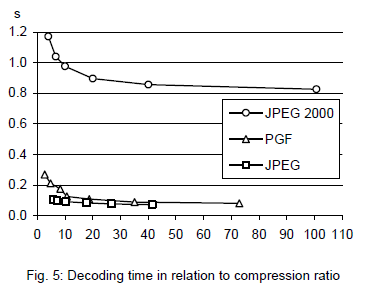


It is very fast and has a reasonably good compression efficiency for a wide range of compression ratios. The drawbacks of JPEG are the missing lossless compression and the often missing progressive decoding. Fig. 4 depicts the average rate-distortion behavior for the images in the Kodak test set when fixed (i.e., nonprogressive) lossy compression is used. The PSNR of PGF is on average 3% smaller than the PSNR of JPEG 2000, but 3% better than JPEG.

These results are also qualitative valid for our PGF test set and they are characteristic for aerial ortho-photos and natural images. Because of the design of PGF we already know that PGF does not reach the compression efficiency of JPEG 2000. However, we are interested in the trade-off between compression efficiency and runtime. To report this trade-off we show in Table 4 a comparison between JPEG 2000 and PGF and in Fig. 5 (on page 8) we show for the same test series as in Fig. 4 the corresponding average decoding times in relation to compression ratios.Table 4 contains for seven different compression ratios (mean values over the compression ratios of the eight images of the Kodak test set) the corresponding average encoding and decoding times in relation to the average PSNR values. In case of PGF the encoding time is always slightly longer than the corresponding decoding time. The reason for that is that the actual encoding phase (cf. Subsection 2.4.2) takes slightly longer than the corresponding decoding phase. For six of seven ratios the PSNR difference between JPEG 2000 and PGF is within 3% of the PSNR of JPEG 2000. Only in the first row is the difference larger (21%), but because a PSNR of 50 corresponds to an almost perfect image quality the large PSNR difference corresponds with an almost undiscoverable visual difference. The price they pay in JPEG 2000 for the 3% more PSNR is very high. The creation of a PGF is five to twenty times faster than the creation of a corresponding JPEG 2000 file, and the decoding of the created PGF is still five to ten times faster than the decoding of the JPEG 2000 file. This gain in speed is remarkable, especially in areas where time is more important than quality, maybe for instance in real-time computation.



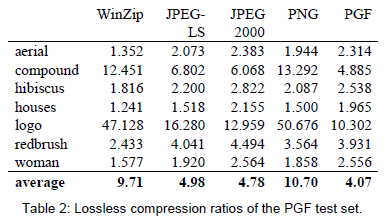
In Fig. 5 we see that the price we pay in PGF for the 3% more PSNR than JPEG is low: for small compression ratios (< 9) decoding in PGF takes two times longer than JPEG and for higher compression ratios (> 30) it takes only ten percent longer than JPEG. These test results are characteristic for both natural images and aerial ortho-photos. Again, in the third test series we only use the ‘Lena’ image. We run our lossy coder with six different quantization parameters and measure the PSNR in relation to the resulting compression ratios. The results (ratio: PSNR) are:



**2.Lossless Image compression :**

Lossless Image compression is the only acceptable amount of data reduction. It provides low compression ratio while compared to lossy. In Lossless Image compression techniques are composed of two relatively independent operations: (1) devising an alternative representation of the image in which its interpixel redundancies are reduced and (2) coding the representation to eliminate coding redundancies.

Lossless Image compression is useful in applications such as medical imaginary, business documents and satellite images.Table 2 summarizes the lossless compression efficiency and Table 3 the coding times of the PGF test set. For WinZip we only provide average runtime values, because of missing source code we have to use an interactive testing procedure with runtimes measured by hand. All other values are measured in batch mode.



In Table 2 it can be seen that in almost all cases the best compression ratio is obtained by JPEG 2000, followed by PGF, JPEG-LS, and PNG. This result is different to the result in [SEA+00], where the best performance for a similar test set has been reported for JPEG-LS. PGF performs between 0.5% (woman) and 21.3% (logo) worse than JPEG 2000. On average it is almost 15% worse. The two exceptions to the general trend are the ‘compound’ and the ‘logo’ images. Both images contain for the most part black text on a white background. For this type of images, JPEG-LS and in particular WinZip and PNG provide much larger compression ratios. However, in average PNG performs the best, which is also reported in [SEA+00].

These results show, that as far as lossless compression is concerned, PGF performs reasonably well on natural and aerial images. In specific types of images such as ‘compound’ and ‘logo’ PGF is outperformed by far in PNG.

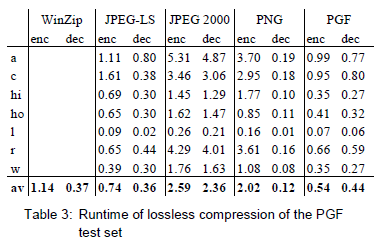


Table 3 shows the encoding (enc) and decoding (dec) times (measured in seconds) for the same algorithms and images as in Table 2. JPEG 2000 and PGF are both symmetric algorithms, while WinZip, JPEG-LS and in particular PNG are asymmetric with a clearly shorter decoding than encoding time. JPEG 2000, the slowest in encoding and decoding, takes more than four times longer than PGF. This speed gain is due to the simpler coding phase of PGF. JPEG-LS is slightly slower than PGF during encoding, but slightly faster in decoding images.

WinZip and PNG decode even more faster than JPEG-LS, but their encoding times are also worse. PGF seems to be the best compromise between encoding and decoding times.

Our PGF test set clearly shows that PGF in lossless mode is best suited for natural images and aerial ortho photos. PGF is the only algorithm that encodes the three Mega Byte large aerial ortho photo in less than second without a real loss of compression efficiency. For this particular image the efficiency loss is less than three percent compared to the best. These results should be underlined with our second test set, the Kodak test set.

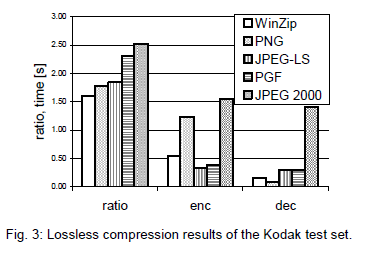


Fig. 3 shows the averages of the compression ratios (ratio), encoding (enc), and decoding (dec) times over all eight images. JPEG 2000 shows in this test set the best compression efficiency followed by PGF, JPEG-LS, PNG, and WinZip. In average PGF is eight percent worse than JPEG 2000. The fact that JPEG 2000 has a better lossless compression ratio than PGF does not surprise,

because JPEG 2000 is more quality driven than PGF.

However, it is remarkable that PGF is clearly better than JPEG-LS (+21%) and PNG (+23%) for natural images. JPEG-LS shows in the Kodak test set also a symmetric encoding and decoding time behaviour. It is encoding and decoding times are almost equal to PGF. Only PNG and WinZip can faster decode than PGF, but they also take longer than PGF to encode.

If both compression efficiency and runtime is important, then PGF is clearly the best of the tested algorithms for lossless compression of natural images and aerial ortho photos. In the third test we perform our lossless coder on the ‘Lena’ image.

To digitally process an image, it is first necessary to reduce the image to a series of numbers that can be manipulated by the computer. Each number representing the brightness value of the image at a particular location is called a picture element, or pixel. A typical digitized image may have 512 × 512 or roughly 250,000 pixels, although much larger images are becoming common. Once the image has been digitized, there are three basic operations that can be performed on it in the computer. For a point operation, a pixel value in the output image depends on a single pixel value in the input image. For local operations, several neighbouring pixels in the input image determine the value of an output image pixel. In a global operation, all of the input image pixels contribute to an output image pixel value.

Correspondingly, these combinations attempt to strike a winning tradeoff: be flexible and hence bring tolerance toward intraclass variation, while also being discriminative enough to be robust to background clutter and interclass similarity. An important feature of our contour-based recognition approach is that it affords us substantial flexibility to incorporate additional image information. Specifically, we extend the contour-based recognition method and propose a new hybrid recognition method which exploits shape tokens and SIFT features as recognition cues. Shape-tokens and SIFT features are largely orthogonal, where the former corresponds to shape boundaries and the latter to sparse salient image patches. Here, each learned combination can comprise features that are either 1) purely shape-tokens, 2) purely SIFT features, or 3) a mixture of shape-tokens and SIFT features. The number and types of features to be combined together are learned automatically from training images, and represent the more discriminative ones based on the training set. Consequently, by imparting these two degrees of variability (in both the number and the types of features) to a combination, we empower it with even greater flexibility and discriminative potential. A shorter version of this paper appeared in [9].

**CLASSIFICATION OF IMAGES:**

There are 3 types of images used in Digital Image Processing. They are

1. Binary Image
2. Gray Scale Image
3. Colour Image

**BINARY IMAGE:**

A binary image is a [digital image](http://en.wikipedia.org/wiki/Digital_image) that has only two possible values for each [pixel](http://en.wikipedia.org/wiki/Pixel).  Typically the two colors used for a binary image are black and white though any two colors can be used.  The color used for the object(s) in the image is the foreground color while the rest of the image is the background color.

Binary images are also called bi-level or two-level. This means that each pixel is stored as a single bit (0 or 1).This name black and white, monochrome or monochromatic are often used for this concept, but may also designate any images that have only one sample per pixel, such as [grayscale images](http://en.wikipedia.org/wiki/Grayscale)

Binary images often arise in [digital image processing](http://en.wikipedia.org/wiki/Digital_image_processing) as [masks](http://en.wikipedia.org/w/index.php?title=Mask_(image_processing)&action=edit&redlink=1) or as the result of certain operations such as [segmentation](http://en.wikipedia.org/wiki/Segmentation_(image_processing)), [thresholding](http://en.wikipedia.org/wiki/Thresholding_(image_processing)), and [dithering](http://en.wikipedia.org/wiki/Dither). Some input/output devices, such as [laser printers](http://en.wikipedia.org/wiki/Laser_printer), [fax machines](http://en.wikipedia.org/wiki/Fax), and bi-level [computer displays](http://en.wikipedia.org/wiki/Visual_display_unit), can only handle bi-level images

**GRAY SCALE IMAGE**

A grayscale Image is [digital image](http://en.wikipedia.org/wiki/Digital_image) is an image in which the value of each [pixel](http://en.wikipedia.org/wiki/Pixel) is a single [sample](http://en.wikipedia.org/wiki/Sample_(signal)), that is, it carries only [intensity](http://en.wikipedia.org/wiki/Luminous_intensity) information. Images of this sort, also known as [black-and-white](http://en.wikipedia.org/wiki/Black-and-white), are composed exclusively of shades of [gray](http://en.wikipedia.org/wiki/Gray" \o "Gray)(0-255), varying from black(0) at the weakest intensity to white(255) at the strongest.

Grayscale images are distinct from one-bit [black-and-white](http://en.wikipedia.org/wiki/Black-and-white) images, which in the context of computer imaging are images with only the two [colors](http://en.wikipedia.org/wiki/Color), [black](http://en.wikipedia.org/wiki/Black), and [white](http://en.wikipedia.org/wiki/White) (also called bi-level or [binary images](http://en.wikipedia.org/wiki/Binary_image)). Grayscale images have many shades of gray in between. Grayscale images are also called [monochromatic](http://en.wikipedia.org/wiki/Monochromatic), denoting the absence of any [chromatic](http://en.wikipedia.org/wiki/Chromaticity) variation.

Grayscale images are often the result of measuring the intensity of light at each pixel in a single band of the [electromagnetic spectrum](http://en.wikipedia.org/wiki/Electromagnetic_spectrum) (e.g. [infrared](http://en.wikipedia.org/wiki/Infrared), [visible light](http://en.wikipedia.org/wiki/Visible_spectrum), [ultraviolet](http://en.wikipedia.org/wiki/Ultraviolet), etc.), and in such cases they are monochromatic proper when only a given [frequency](http://en.wikipedia.org/wiki/Frequency) is captured. But also they can be synthesized from a full color image; see the section about converting to grayscale.

**COLOUR IMAGE:**

A (digital) color image is a [digital image](http://en.wikipedia.org/wiki/Digital_image) that includes [color](http://en.wikipedia.org/wiki/Color) information for each [pixel](http://en.wikipedia.org/wiki/Pixel). Each pixel has a particular value which determines its appearing color. This value is qualified by three numbers giving the decomposition of the color in the three primary colors Red, Green and Blue. Any color visible to human eye can be represented this way. The decomposition of a color in the three primary colors is quantified by a number between 0 and 255. For example, white will be coded as R = 255, G = 255, B = 255; black will be known as (R,G,B) = (0,0,0); and say, bright pink will be : (255,0,255).

In other words, an image is an enormous two-dimensional array of color values, pixels, each of them coded on 3 bytes, representing the three primary colors. This allows the image to contain a total of 256x256x256 = 16.8 million different colors. This technique is also known as RGB encoding, and is specifically adapted to human vision

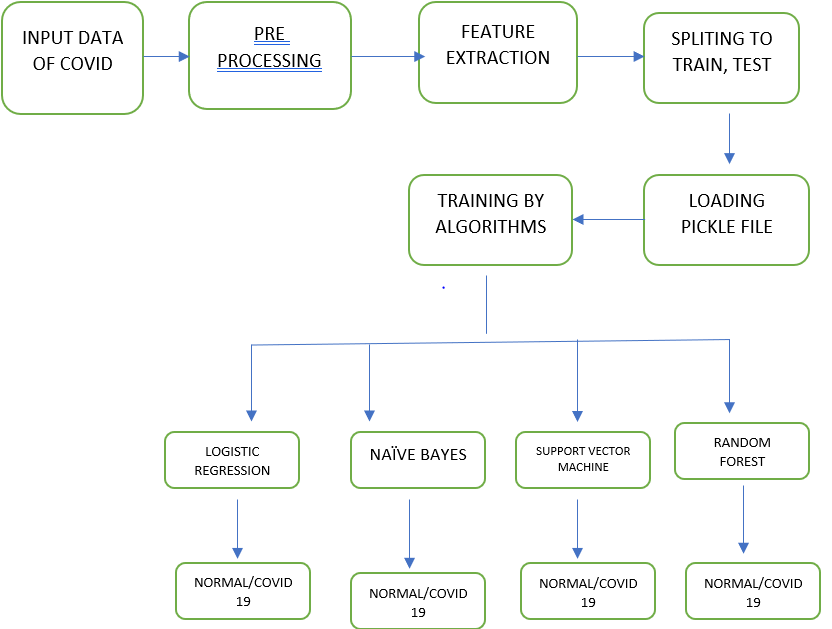
|  |
| --- |
| http://images.gamedev.net/features/programming/imageproc/image004.gif |

**Fig.1 Hue Saturation Process of RGB SCALE Image**

From the above figure, colors are coded on three bytes representing their decomposition on the three primary colors. It sounds obvious to a mathematician to immediately interpret colors as vectors in a three dimension space where each axis stands for one of the primary colors. Therefore we will benefit of most of the geometric mathematical concepts to deal with our colors, such as norms, scalar product, projection, rotation or distance.

3.PROPOSED METHODOLOGY:

BLOCK DIAGRAM



The dataset used and the methodology used is explained in the subsequent sections

3.1 DATASET:

The dataset of this work has been collected from Kaggle reposivory , that contains Chest X-Ray scans of Covid-19 affected, traditional and respiratory illness. This collected dataset isn't meant to say the diagnostic ability of any Deep Learning model however to analysis regarding numerous attainable ways that of with efficiency detecting Coronavirus infections victimization laptop vision techniques. The collected dataset consists of 6432 total chest X-ray pictures. This information set is more divided into coaching (i.e., 5467) and validation (i.e., 965) set of traditional, covid, and respiratory illness. In the training set, 1345 is traditional, 490 ar covid, and 3632 is pneumonia. within the validation section, 238 samples of a traditional case, 86 covid, and 641 of respiratory illness were thought-about for this analysis. At the time of the drafting of this paper, we have a tendency to had 576 PA read scans of Covid-19 affected patients. The scans were scaled down 128 × 128 to help the quick coaching of our model. The PA read scans were deemed to be in keeping with our covid dataset. Table one displays the info distribution for training and testing the info.

3.2 MODEL FORMULATION

The data obtained from the Kaggle repository was clean as needed. To implement a deep learning methodology needs a large quantity of dataset to receive reliable results. however it might be attainable that each downside doesn't have enough information, especially in medical-related problems. typically grouping medical-related information is also time overwhelming and big-ticket. To solve these types of difficulties, augmentation are often applied. Augmentation will overcome the matter of over-fitting and enhance the accuracy of the planned model. Further, augmentation is applied during this collected dataset to prevent over-fitting. The augmentations enclosed rotation, zoom, and sharing of pictures. the info was then shuffled to generalize the model and scale back over-fitting. After this, the ready dataset was accustomed train the planned model. For higher analysis, three different models are enforced, so their performance was compared to calculate the accuracy. within the given models, we tend to enforced pickle activation rather than the originally used rely activation perform, that makes it as a completely unique method. This method helps to hurry up the coaching and conjointly avoids the matter of dead neurons the planned model for chest x-ray image analysis.

3.3 PROPOSED ALGORITHMS:

LOGISTIC REGRESSION:

It is accustomed estimate separate values ( Binary values like 0/1, yes/no, true/false ) supported given set of freelance variable(s). In easy words, it predicts the likelihood of incidence of an occurrence by fitting information to a logit operate. Hence, it's conjointly called logit regression. Since, it predicts the likelihood, its output values lies between zero and one (as expected).

NAÏVE BAYES:

It is a classification technique supported Bayes’ theorem with associate degree assumption of independence between predictors. In straightforward terms, a Naive mathematician categorified assumes that the presence of a specific feature during a class is unrelated to the presence of the other feature. for instance, a fruit could also be thought of to be associate degree apple if it's red, round, and regarding three inches in diameter. albeit these options rely on one another or upon the existence of the opposite options, a naive mathematician classifier would think about all of those properties to severally contribute to the chance that this fruit is associate degree apple.

SUPPORT VECOR MCAHINE:

In this algorithmic rule, we have a tendency to plot every knowledge item as a degree in n-dimensional house (where n is variety of options you have) with the worth of every feature being the worth of a specific coordinate.(these co-ordinates square measure referred to as Support Vectors)

RANDOM FOREST:

Random Forest could be a trademark term for Associate in Nursing ensemble of call trees. In Random Forest, we’ve assortment of call trees (so referred to as “Forest”). To classify a brand new object supported attributes, every tree offers a categorification and that we say the tree “votes” for that class. The forest chooses the classification having the foremost votes (over all the trees within the forest).

Matplotlib

Matplotlib is a plotting library for Python. It is used along with NumPy to provide an environment that is an effective open source alternative for MatLab. It can also be used with graphics toolkits like PyQt and wxPython.

Matplotlib module was first written by John D. Hunter. Since 2012, Michael Droettboom is the principal developer. Currently, Matplotlib ver. 1.5.1 is the stable version available. The package is available in binary distribution as well as in the source code form on [www.matplotlib.org](http://www.matplotlib.org/).

Conventionally, the package is imported into the Python script by adding the following statement −

from matplotlib import pyplot as plt

Here **pyplot()** is the most important function in matplotlib library, which is used to plot 2D data. The following script plots the equation **y = 2x + 5**

An ndarray object x is created from **np.arange() function** as the values on the **x axis**. The corresponding values on the **y axis** are stored in another **ndarray object y**. These values are plotted using **plot()** function of pyplot submodule of matplotlib package.

The graphical representation is displayed by **show()** function.

Numpy

NumPy, which stands for Numerical Python, is a library consisting of multidimensional array objects and a collection of routines for processing those arrays. Using NumPy, mathematical and logical operations on arrays can be performed. This tutorial explains the basics of NumPy such as its architecture and environment. It also discusses the various array functions, types of indexing, etc. An introduction to Matplotlib is also provided. All this is explained with the help of examples for better understanding.

Audience

This tutorial has been prepared for those who want to learn about the basics and various functions of NumPy. It is specifically useful for algorithm developers. After completing this tutorial, you will find yourself at a moderate level of expertise from where you can take yourself to higher levels of expertise.

Prerequisites

You should have a basic understanding of computer programming terminologies. A basic understanding of Python and any of the programming languages is a plus.

NumPy is a Python package. It stands for 'Numerical Python'. It is a library consisting of multidimensional array objects and a collection of routines for processing of array.

**Numeric**, the ancestor of NumPy, was developed by Jim Hugunin. Another package Numarray was also developed, having some additional functionalities. In 2005, Travis Oliphant created NumPy package by incorporating the features of Numarray into Numeric package. There are many contributors to this open source project.

## Operations using NumPy

Using NumPy, a developer can perform the following operations −

* Mathematical and logical operations on arrays.
* Fourier transforms and routines for shape manipulation.
* Operations related to linear algebra. NumPy has in-built functions for linear algebra and random number generation.

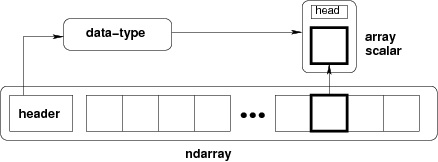
## NumPy – A Replacement for MatLab

NumPy is often used along with packages like **SciPy** (Scientific Python) and **Mat−plotlib** (plotting library). This combination is widely used as a replacement for MatLab, a popular platform for technical computing. However, Python alternative to MatLab is now seen as a more modern and complete programming language.

The most important object defined in NumPy is an N-dimensional array type called **ndarray**. It describes the collection of items of the same type. Items in the collection can be accessed using a zero-based index.

Every item in an ndarray takes the same size of block in the memory. Each element in ndarray is an object of data-type object (called **dtype**).

Any item extracted from ndarray object (by slicing) is represented by a Python object of one of array scalar types. The following diagram shows a relationship between ndarray, data type object (dtype) and array scalar type −



An instance of ndarray class can be constructed by different array creation routines described later in the tutorial. The basic ndarray is created using an array function in NumPy as follows −

numpy.array

It creates an ndarray from any object exposing array interface, or from any method that returns an array.

Software used

1.What is Python

**Python** is an object-oriented, high level language, interpreted, dynamic and multipurpose programming language.

Python is *easy to learn* yet powerful and versatile scripting language which makes it attractive for Application Development.

Python's syntax and *dynamic typing* with its interpreted nature, make it an ideal language for scripting and rapid application development in many areas.

Python supports *multiple programming pattern*, including object oriented programming, imperative and functional programming or procedural styles.

Python is not intended to work on special area such as web programming. That is why it is known as *multipurpose* because it can be used with web, enterprise, 3D CAD etc.

We don't need to use data types to declare variable because it is *dynamically typed* so we can write a=10 to declare an integer value in a variable.

Python makes the development and debugging *fast* because there is no compilation step included in python development and edit-test-debug cycle is very fast.

2. Python Features

#### 1) Easy to Use:

Python is easy to very easy to use and high level language. Thus it is programmer-friendly language.

#### 2) Expressive Language:

Python language is more expressive. The sense of expressive is the code is easily understandable.

#### 3) Interpreted Language:

Python is an interpreted language i.e. interpreter executes the code line by line at a time. This makes debugging easy and thus suitable for beginners.

#### 4) Cross-platform language:

Python can run equally on different platforms such as Windows, Linux, Unix , Macintosh etc. Thus, Python is a portable language.

#### 5) Free and Open Source:

Python language is freely available(www.python.org).The source-code is also available. Therefore it is open source.

#### 6) Object-Oriented language:

Python supports object oriented language. Concept of classes and objects comes into existence.

#### 7) Extensible:

It implies that other languages such as C/C++ can be used to compile the code and thus it can be used further in your python code.

#### 8) Large Standard Library:

Python has a large and broad library.

#### 9) GUI Programming:

Graphical user interfaces can be developed using Python.

#### 10) Integrated:

It can be easily integrated with languages like C, C++, JAVA etc.

3. Python History

* Python laid its foundation in the late 1980s.
* The implementation of Python was started in the December 1989 by **Guido Van Rossum** at CWI in Netherland.
* *ABC programming language* is said to be the predecessor of Python language which was capable of Exception Handling and interfacing with Amoeba Operating System.
* Python is influenced by programming languages like:
  + ABC language.
  + Modula-3

# 4. Python Version

Python programming language is being updated regularly with new features and support. There are a lot of updation in python versions, started from 1994 to current date.

A list of python versions with its released date is given below.

|  |  |
| --- | --- |
| **Python Version** | **Released Date** |
| Python 1.0 | January 1994 |
| Python 1.5 | December 31, 1997 |
| Python 1.6 | September 5, 2000 |
| Python 2.0 | October 16, 2000 |
| Python 2.1 | April 17, 2001 |
| Python 2.2 | December 21, 2001 |
| Python 2.3 | July 29, 2003 |
| Python 2.4 | November 30, 2004 |
| Python 2.5 | September 19, 2006 |
| Python 2.6 | October 1, 2008 |
| Python 2.7 | July 3, 2010 |
| Python 3.0 | December 3, 2008 |
| Python 3.1 | June 27, 2009 |
| Python 3.2 | February 20, 2011 |
| Python 3.3 | September 29, 2012 |

# 5. Python Applications

Python as a whole can be used in any sphere of development.

Let us see what are the major regions where Python proves to be handy.

#### 1) Console Based Application

Python can be used to develop console based applications. For example: **IPython**.

#### 2) Audio or Video based Applications

Python proves handy in multimedia section. Some of real applications are: TimPlayer, cplay etc.

#### 3) 3D CAD Applications

Fandango is a real application which provides full features of CAD.

#### 4) Web Applications

Python can also be used to develop web based application. Some important developments are: PythonWikiEngines, Pocoo, PythonBlogSoftware etc.

#### 5) Enterprise Applications

Python can be used to create applications which can be used within an Enterprise or an Organization. Some real time applications are: OpenErp, Tryton, Picalo etc.

#### 6) Applications for Images

Using Python several application can be developed for image. Applications developed are: VPython, Gogh, imgSeek etc.

There are several such applications which can be developed using Python

# 6. Python Example

Python code is simple and easy to run. Here is a simple Python code that will print "Welcome to Python".

A simple python example is given below.

1. >>> a="Welcome To Python"
2. >>> **print** a
3. Welcome To Python
4. >>>

**Explanation:**

* Here we are using IDLE to write the Python code. Detail explanation to run code is given in Execute Python section.
* A variable is defined named "a" which holds "Welcome To Python".
* "print" statement is used to print the content. Therefore "print a" statement will print the content of the variable. Therefore, the output "Welcome To Python" is produced.

## Python 3.4 Example

In python 3.4 version, you need to add parenthesis () in a string code to print it.

1. >>> a=("Welcome To Python Example")
2. >>> **print** a
3. Welcome To Python Example
4. >>>

# 7. How to execute python

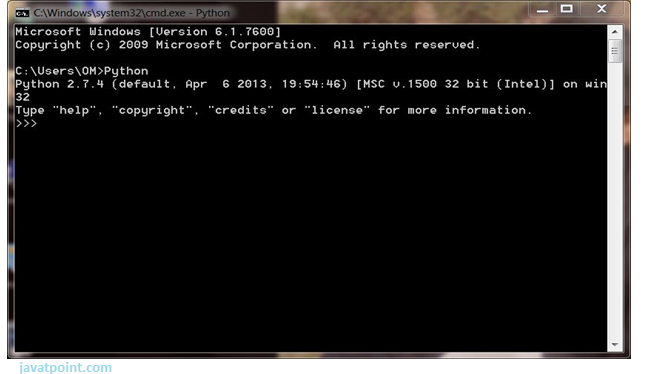
There are three different ways of working in Python:

## 1) Interactive Mode:

You can enter python in the command prompt and start working with Python.

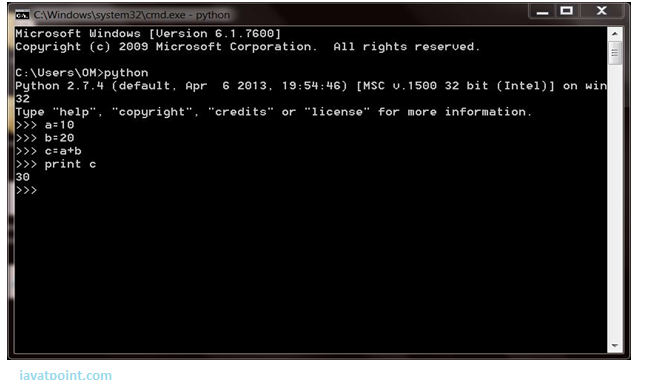


Press Enter key and the Command Prompt will appear like:



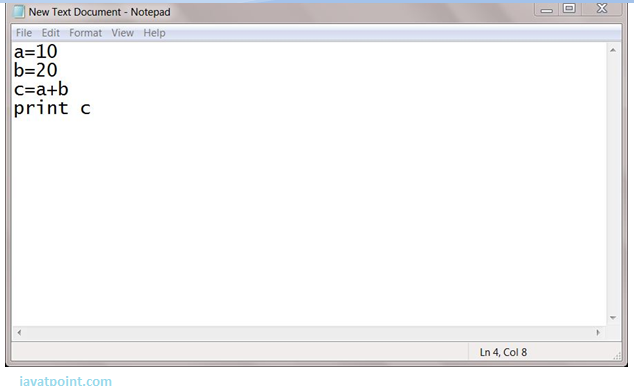
Now you can execute your Python commands.

**Eg:**

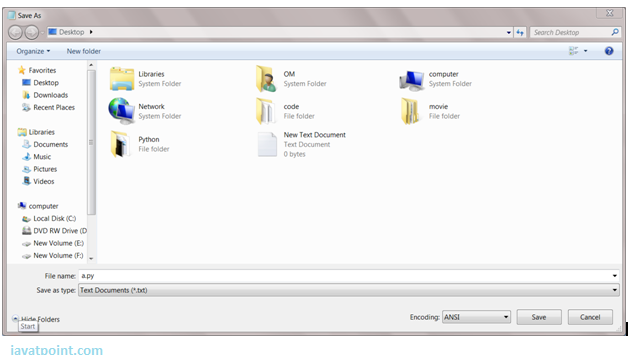


## 2) Script Mode:

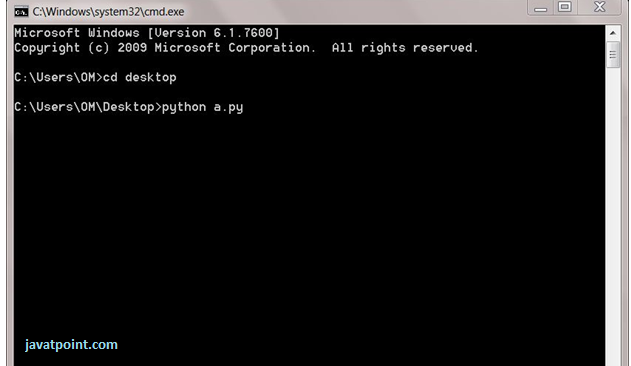
Using Script Mode , you can write your Python code in a separate file using any editor of your Operating System.



Save it by .py extension.



Now open Command prompt and execute it by :



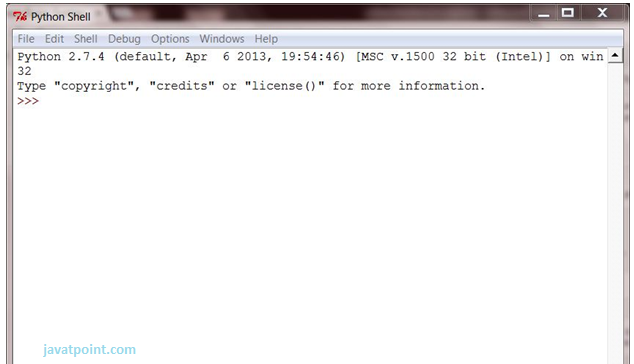
NOTE: Path in the command prompt should be where you have saved your file. In the above case file should be saved at desktop.

## 3) Using IDE: (Integrated Development Environment)

You can execute your Python code using a Graphical User Interface (GUI).

All you need to do is:

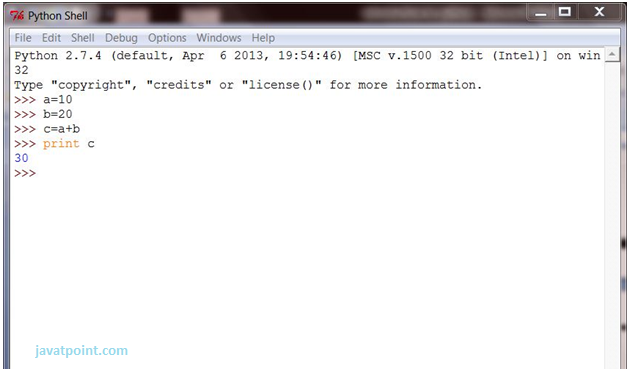
Click on Start button -> All Programs -> Python -> IDLE(Python GUI)



You can use both Interactive as well as Script mode in IDE.

**1) Using Interactive mode:**

Execute your Python code on the Python prompt and it will display result simultaneously.

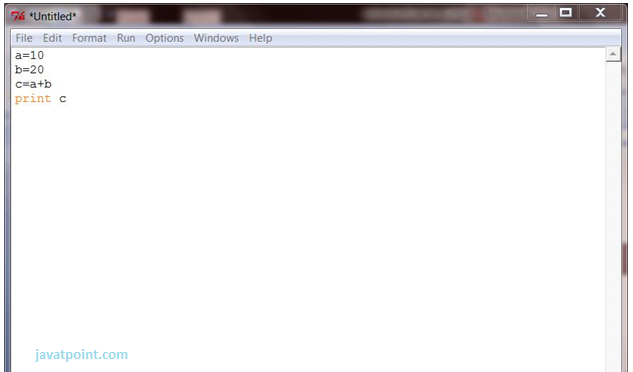


**2) Using Script Mode:**

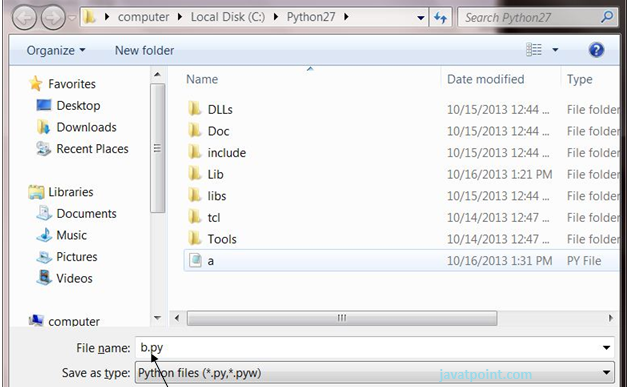
i) Click on Start button -> All Programs -> Python -> IDLE(Python GUI)

ii) Python Shell will be opened. Now click on File -> New Window.

A new Editor will be opened . Write your Python code here.



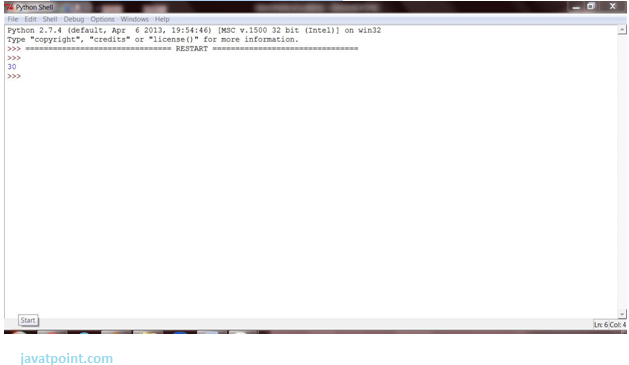
Click on file -> save as



Run then code by clicking on Run in the Menu bar.

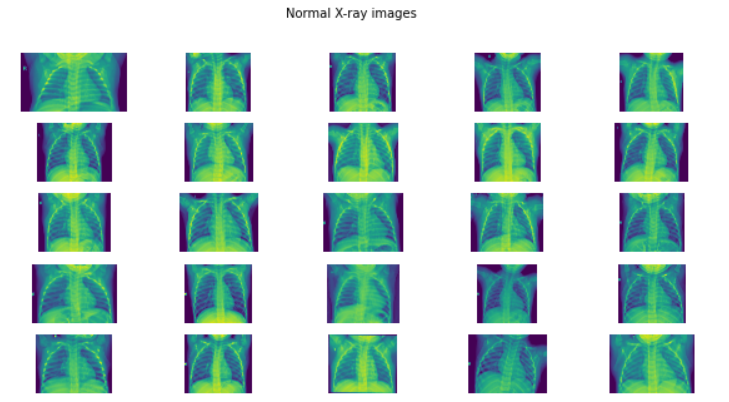
Run -> Run Module

Result will be displayed on a new Python shell as:



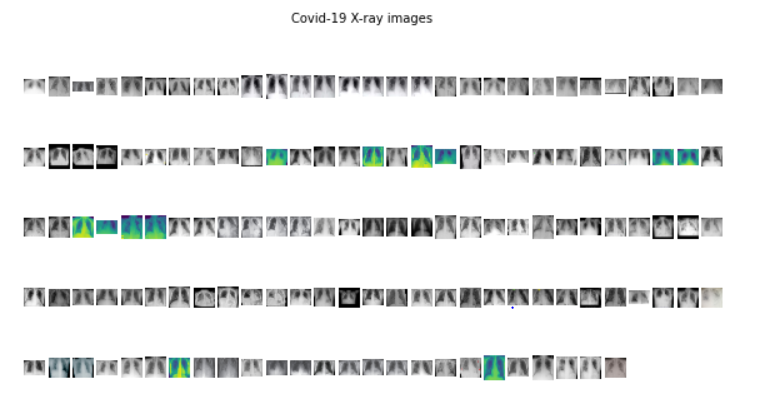
RESULTS AND DISCUSSION:

RESULT1



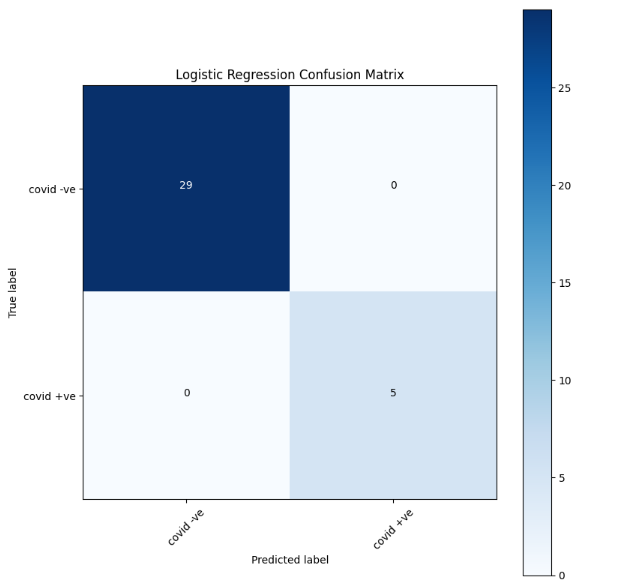
Plotting normal x -ray images from the dataset using matplotlib module

RESULT2



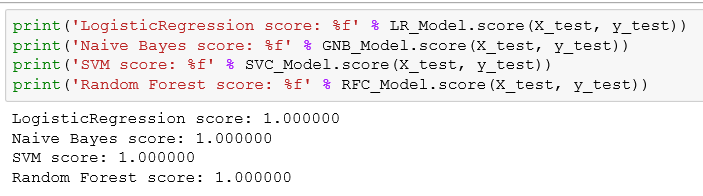
Plotting covid -19 x -ray images from the dataset using matplotlib module

RESULT3:

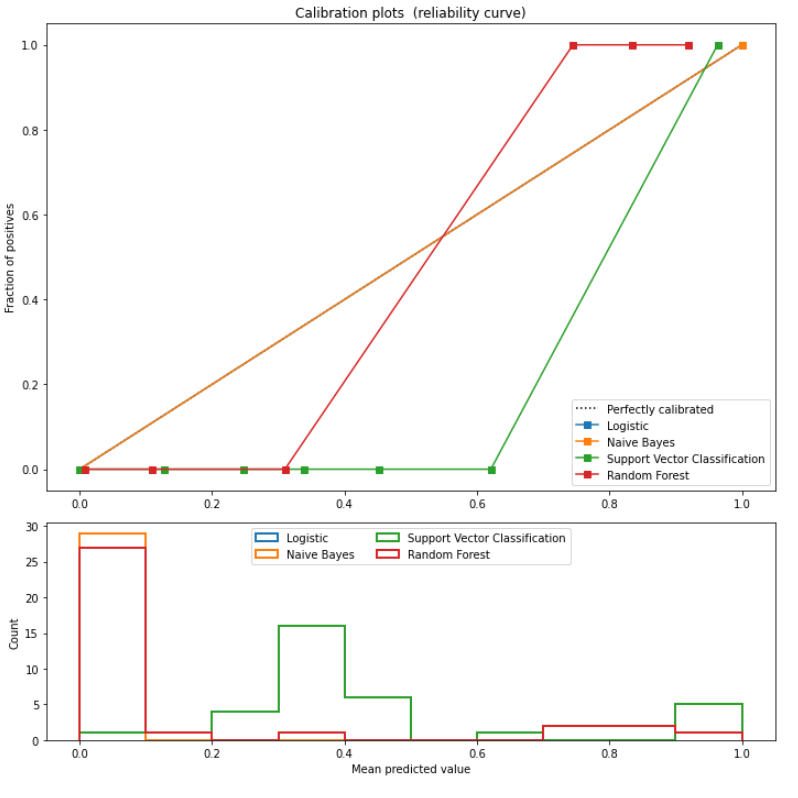


Plotting confusion matrix using logistic regression algorithm

RESULT4:

 Finding Same accuracy for all the algorithms

RESULT5:



Predicted values comparison for four algorithms

CONCLUSION:

COVID-19 has appalled the planet thanks to its non-availability of vaccinum or drug. varied researchers square measure operating for subjugation this deadly virus. varied options like TF/IDF, bag of words square measure being extracted from these clinical reports. The machine learning algorithms square measure used for classifying clinical reports into four totally different categories. when playacting classification, it absolutely was unconcealed that supplying regression and multinomial Naïve Bayesian classifier provides wonderful results by having ninety four preciseness, 96% recall, ninety fifth f1 score and accuracy ninety six.2%. varied alternative machine learning algorithms that showed higher results were random forest, random gradient boosting, call trees and boosting. The potency of models may be improved by increasing the number of information. Also, the sickness may be classified on the gender-based such get info concerning whether or not male are affected additional or females. additional feature engineering is required for higher results and deep learning approach may be employed in future.

REREFERENCES:

[1] Roosa K, Lee Y, Luo R, Kirpich A, Rothenberg R, Hyman JM, and et al. Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020. Infectious Disease Modelling, 5:256-263, 2020.

[2] Yan L, Zhang H-T, Xiao Y, Wang M, Sun C, Liang J, and et al. Prediction of criticality in patients with severe COVID-19 infection using three clinical features: a machine learningbased prognostic model with clinical data in Wuhan. medRxiv 2020.02.27.20028027, 2020.

[3] Stoecklin SB, Rolland P, Silue Y, Mailles A, Campese C, Simondon A, and et al. First cases of coronavirus disease 2019 (COVID-19) in France: surveillance, investigations and control measures, January 2020. Eurosurveillance, 25(6):2000094, 2020.

[4] Coronavirus. World Health Organization: https://www.who.int/healthtopics/coronavirus, 2020.

[5] McKeever A. Here's what coronavirus does to the body. National Geographic: https://www.nationalgeographic.com/science/2020/02/here-is-what-coronavirus-doesto-thebody/, 2020.

[6] COVID-19 Coronavirus Pandemic. worldometer: https://www.worldometers.info/ coronavirus/, 2020.

[7] Jaiswal AK, Tiwari P, Kumar S, Gupta D, Khanna A, and Rodrigues JJ. Identifying pneumonia in chest X-rays: A deep learning approach. Measurement, 145:511-518, 2019.

[8] Antin B, Kravitz J, and Martayan E. Detecting Pneumonia in Chest X-Rays with Supervised Learning. http://cs229.stanford.edu/proj2017/final-reports/5231221.pdf, 1-5, 2017.

[9] Das NN, Kumar N, Kaur M, Kumar V, and Singh D. Automated Deep Transfer LearningBased Approach for Detection of COVID-19 Infection in Chest X-rays. IRBM, https://doi.org/10.1016/j.irbm.2020.07.001, 2020