A Fungus Spores Dataset and a Convolutional Neural Networks based Approach for Fungus Detection

Muhammad Waseem Tahir, Nayyer Abbas Zaidi, Adeel Akhtar Rao, Roland Blank, Michael J. Vellekoop and Walter Lang

***Abstract*— Fungus is an enormously notorious for food, human health and archives. Fungus sign and symptoms in medical science are non-specific and asymmetrical for extremely large areas resulting into a challenging task of fungal detection. Various traditional and computer vision techniques were applied to meet the challenge of early fungus detection. On the other hand, features learned through the convolutional neural network (CNN) provided state of the art results in many other applications of object detection and classification. However, the large amount of data is an essential prerequisite for its effective application. In pursuing this idea, we present a novel fungus dataset of its kind, with the goal of advancing the state-of-the-art in fungus classification by placing the question of fungus detection. This is achieved by gathering various images of complex fungal spores by extracting samples from contaminated fruits, archives and lab incubated fungus colonies. These images primarily consisted of five different types of fungus spores and dirt. Optical sensor system was utilized to obtain these images. Which were further annotated to mark fungal spores as a region of interest using specially designed graphical user interface. As a result, 40,800 labeled images were used to develop fungus dataset to aid in precise fungus detection and classification. The other main objective of this research was to develop a CNN based approach for the detection of fungus and distinguish different types of fungus. A CNN architecture was designed and it showed the promising results with an accuracy of 94.8%. The obtained results proved the possibility of early detection of several types of fungus spores using CNN and could estimate all possible threats due to fungus.**

***Index Terms*— Convolutional neural network, deep learning, fungus detection, fungus dataset, microbial detection, mold detection, mold dataset.**

Manuscript received February 12, 2018, and revised on March 25, 2018, April 20, 2018. This work was supported in parts by DAAD/HEC and the Federal Ministry of Education and Research, Germany, through Project MaUS: Mikroreaktorensystem zur autonomen Untersuchung von Schimmelpilzbelastungen, Project No 031A257A-D.

M. W. Tahir and N. A. Zaidi work at the Institute for Microsensors, -Actuators and -Systems (IMSAS), University of Bremen, 28359 Bremen, Germany. They are also affiliated with International Graduate School for Dynamics in Logistics (IGS), University of Bremen, Germany.(email: wtahir@imsas.uni-bremen.de; nzaidi@imsas.uni-bremen.de

A. A. Rao works in KPMG AG WPG at data and Analytics department in Dsseldorf.(email: adeelrao90@gmail.com).

R. Blank, M. J. Vellekoop and W. Lang are with the Institute for Microsensors, -Actuators and -Systems, University of Bremen, Bremen 28359, Germany (email: rblank@imsas.uni-bremen.de; mvellekoop@imsas. uni-bremen.de; wlang@imsas.uni-bremen.de).

I. Introduction

FUNGUS is an essential eukaryotic microorganism of the ecosystem, which plays an important role not just in the decomposition process of organic matter with its more than 100,000 unique recognized species but also enormously notorious for human health, food and archives for making them prone to its threat. The fungus can affect human health in various ways, which include systemic infection to human tissues, respiration problem, decay of stored food in containers with possible production of mycotoxins and biodeterioration of library material (e.g. books and documents). Fungi pose a potential hazard to agriculture as well. Half of the harvested fruits and vegetables have been lost due to fungal infection in tropics [1]. Moreover, one-third of the yielded food is frittered away each year [2]. In this modern era, actions should be performed for early detection and elimination of fungus. To guard and secure public health and food items, we need to carry out continuous screening of fungus and its spores presence in our environment.

There exist several methods to detect fungal diseases like [3]–[6]. Traditional methods included plate count method, nucleic acid probe technology and polymerase chain reaction (PCR) [7]. Some of these methods require special equipment as well as well-trained labor to perform the task. Scientists carried out sophisticated analysis, usually with the help of powerful microscopes. Similarly, other researchers also used electromagnetic spectrum, which is not visible by humans, helps to detect the fungal contamination. Accurate and reliable results were obtained on the cost of expensive equipment and expert labor was trusted to screen fungal spores that required huge processing time. In the meantime, fungus grows in enormous amount. So for a quick, accurate and spontaneous result permits us to look for the computer vision and machine learning based solution for this problem.

In the last decade, scientists also modernized the process and used computer vision for the ease. Blasco et al. [8] presented an approach for the quality control. Mehrten et al. [9] utilized it for the product classification. and Tahir et al. [10] used it for the detection of fungus. This is one of the reasons that food industry is among the top ten industries that

utilizes computer vision [11].

Before setting the main goal of this research, one important question of concern is to detect fungal hyphae or fungus spores for timely estimation of fungus threat. Fungal hyphae are easier to detect as compare to fungus spores but it is difficult to take timely precautionary measures once these fungus hyphae have been already grown. Where as fungus spores timely detection enable us to control the further growth of fungus spores into fungal hyphae. Hence buy us enough time to regulate environmental conditions. Furthermore, for spore sample collection, a cumbersome manual procedure is not required as they are more resistant to unfavourable environmental conditions. Whereas for fungal hyphae, the sample preparation requires suitable culture media and environmental conditions, which makes this task more difficult and time consuming. Therefore, detection of fungal hyphae will not serve the goal of this research. As they have already been penetrated into substrate and damaged the human life, animals, food and archives. So in this research, we presented a meaningful and useful way, to tackle the difficult problem of fungus control. It is difficult to detect half and fully occluded fungus spores. Illumination changes are also a problem. To solve this problem, we have developed a huge dataset of five different types fungus spores using three different type of light sources. Moreover, deep learning, state of the art technology, based technique applied to achieve the desired goal of fungus detection and a CNN architecture is presented.

The rest of the paper is organized as follows. Section II briefly covers the prior techniques for the fungus detection. Section III provides the detailed description of image collection. The complete methodology used for the fungus detection has been discussed in section IV. Section V presents experimental results and discussion. And at last, Section VI provides the conclusion.

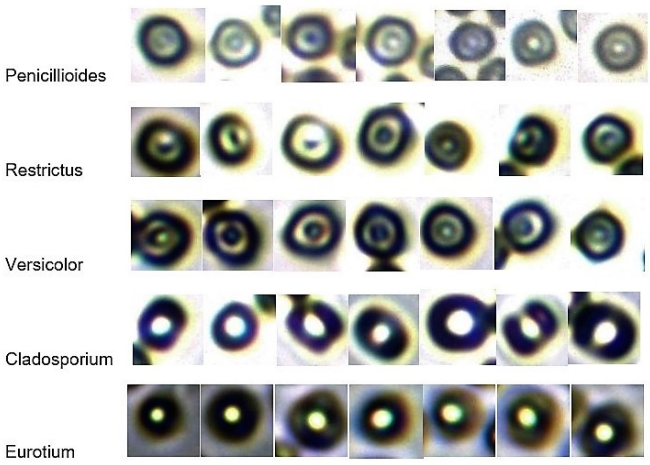


Fig. 1: Overview of Novel fungus dataset

II. Background and Related Work

In last decade, plenty of work was performed in the field of deep learning. In this section, we briefly concise the publicly available computer vision datasets, CNN architectures and latest computer vision techniques used for the detection of microbial.

A. Computer Vision Dataset

Datasets have played a critical role throughout the history of computer vision and image processing. This drives research in various and challenging directions as they provide a meaningful way to train and evaluate algorithms. Prominent advancement has been in some computer vision issues because of a solid competition empowered by freely accessible datasets and evaluation metrics for benchmarking, for example, UCF 50 [12] and UCF 101 [13] for action recognition problem, CIFAR 10 [14], CIFAR 100 [15] and ImageNet [16] for the object detection, object localization and object classification problems. Recently medical imaging also follows this trend and have introduced many datasets and held competitions to stimulate a flood of interest in the field of segmentation, detection, and classification.

For nuclei (that look like fungus spores) segmentation and detection, a large dataset of 29,000 images [17] with marked centers was released but annotations of nuclear boundaries were missing for training and testing. But in [18], Wienert et al. presented a dataset of microscopic images for cellular and tissue specimens with annotation of boundaries. Similarly, Gelasca et al. [19] produced benchmark dataset for image segmentation and tracking. It consists of different classes of images and ground truth data of subcellular, cellular and tissue level.

B. CNN Architectures

Image processing techniques often do not deliver state of the art results in the field of biological sciences like the fungal disease. Therefore, classification methods provide the extension to detect problems with the help of well-known CNN. Several architectures have contributed immensely to the field of convolutional neural network. The researchers had applied numerous techniques regarding CNN to perform classification.

LeCunn et al [20] developed LeNet, the best successful application of CNN without graphical processing units in 1998. LeNet is the best-known architecture of late 90's for recognizing the digits and zip codes. The work popularized in past years for convolution neural network in the computer vision was AlexNet [21], developed by Krizhevsky et al. Submission of AlexNet was done in ImageNet ILSVRC Challenge 2012 by outperforming the second runner-up team. This specific network had similar architecture to Lenet, but consisted of bigger, deeper and featured convolutional layers arranged on top of each other. Initially, it consisted of only single convolution layer which instantly followed by pool

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNB.2018.2839585, IEEE Transactions on NanoBioscience

layer.

Zeiler et al [22] implemented a new convolution network architecture called ZFNet and won ILSVRC 2013 challenge. They provided a new advancement on AlexNet by increasing the size of the middle convolutional layer, changing architecture hyperparameters, applying smaller stride and filter size on the first layer. Later on, Szegedy et al [23] from Google won the ILSVRC 2014 Challenge by implementing convolution network called GoogleNet. The main objective of GoogleNet was to introduce an Inception Module which dramatically reduced the number of parameters in the network. Moreover, this approach eliminated plenty of parameters which were not necessary and utilized average pooling layer rather than the fully connected layer at the top of convolution network.

Simonyan et al [24] became the runner-up in ILSVRC 2014 Challenge by using convolution network and became known as VGGNet. The main part of this network was to introduce that the depth of the convolution network, which is an important tool for enhancing the performance. Their network featured the homogeneous architecture with 2x2 pooling and 3x3 convolution from start to end and consisted of 16 fully connected/convolution layer. This network is more expensive to implement which is the downside of their respective network and uses a lot of parameters sharing over a large memory. A large number of parameters exists in a first fully connected layer. It was since analyzed that these fully connected parameters can be removed without any performance degradation and further reduced the necessary number of parameters.

Kaiman He et al [25] won the ILSVRC 2015 Challenge and developed a residual network called ResNet. Heavy implementation of batch normalization and significant skip connections were introduced in his network. The architecture also refrained itself from using fully connected layer at the end of the network. For implementing convolution neural network, ResNets are the current state of the art for respective neural network and the current choice for implementing neural networks.

This completed the overview of well-known architectures of deep learning. The following paragraph provides the overview of the recent research articles which were focused on using of CNN for the detection and classification tasks of microbial.

Dong, et al. [26] proposed a unique framework for automatic detecting Tyrosine hydroxylase (TH) containing cells in larval zebra fish brain z-stack images recorded through a wide field microscope. In this framework, a supervised max-pool convolution neural network (CNN) was trained to detect cells pixels in regions that were preselected by support vector machine (SVM) classifier. Meanwhile, Chen et al. [27] suggested a novel method for automatic immune cell counting on Immunohistochemistry (IHC) stained slides. A sparse color unmixing technique was utilized to separate the IHC image into multiple color channel corresponding to different cell structure. In close relation to previous research, Malon et al. [28] presented a system to identify all *mitotic* figures in a region of interest of *hematoxylin* and *eosin* stained tissues. This approach utilized convolution neural network for the extraction of learned features with manually designed nuclear features. The nuclear feature captured texture, color and shape information of segmented regions around a nucleus.

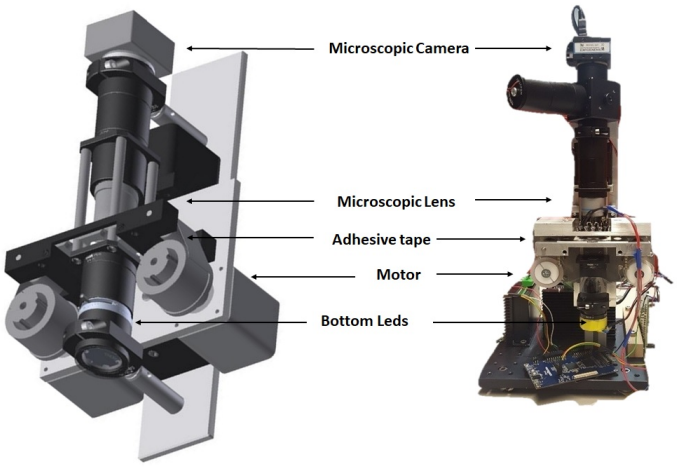


Fig. 2: Image acquisition setup

III. Image Collection

We next describe, which fungus categories were targeted in this research and how images were collected. How annotations of images were done and the setup of our system is also elaborated in this section.

A. Fungus Categories

A thorough study was done regarding different types of fungus infection. The study includes information about the origin, nomenclature, colonies, the occurrence of respective types of fungus. As fungi are microbes and member of the group of eukaryotic organisms which consists of unicellular micro organisms. These organisms are classified as a kingdom. Fungi separate itself from other eukaryotic life kingdoms of animals and plants. Estimated statistics results that there exists from 700,000 to 5 million species of fungi in the world. But only 5 abundantly available fungal species of fruits, archives and environment were primarily focus of our dataset. How they look like in presence of three different light sources (autofluorescence, dark field, bright field) are shown in Fig. 3.

Penicillioides, Restrictus, Versicolor, Cladosporium and Eurotium samples were gathered for this research. These respective fungi were cultured under the supervision of qualified biologist providing the corresponding environment in order to maintain them in their original condition. These fungus samples were cultured at IMSAS Lab in Petri dishes and placed in fungus incubator, where they grow their colonies. After

preparing the samples of respective fungus, these samples were processed under image acquisition system as shown in Fig. 2.

B. Air sampling unit

Air plays a central role as a reservoir for microorganisms, so a special focus has been placed on microbial air sampling. An "Air Sampling Unit (MBASS30)" (ASU) [29] together with Petri dishes containing the samples of fungus and an air pump placed in an airtight container. As the fungus is capable of spreading and growing while dispersing its spores in the air naturally. As expected, these spores were lifted in the air with the help of air pump as a blower. ASU with its PS 30 head, it can generate up to three samples per slide, was utilized for collecting these fungus spores from the air. These samples further investigated using the optical sensor system.

C. Image Acquisition Setup

An optical sensor system as shown in Fig. 2 was used for the important task of image acquisition. Its earlier versions were also presented in [10], [30], [31] and the embedded system of this system is presented in [32], in light of the basic idea of complete fungus detection system as presented in [33]. The imaging source 5 megapixel camera (DFK 72AUC02) was utilized in it. This industrial camera has rolling shutter and 1/2.5 inch Micron CMOS sensor(MT9P031). It can obtain images at 6 fps and pixel size of 2.2x2.2 µm. A Nikon 40X microscopic lens is contacted in front of the camera to obtain high-resolution images. This system has a platform for placing fungus slides, which can move in X, Y and Z axes according to the required operation. This operation is controlled by the external knob as a motor controller, which is capable of its movement. The camera is attached to a laptop, in order to analyze different fungus images when placed under observation. IC Capture 2.4 was used to tune numerous parameters for the enhanced images of fungus. Following are the main points which were kept in consideration.

Parameters tuning: There are different parameters related to tuning such as gain, exposure, gamma and white red balance gives specific enhancement. For each lightning source, these parameters are very important, as they differentiate each class from another. This task requires a lot of concentration, as a minute difference in value changes the nomenclature of fungus images within their domain. Numerous readings were taken, for each image of every fungus type. These parameters are prominent to intensity variation and differ when room light changes. So samples images were collected at daylight timings in order to prevent the errors regarding specific values assigned to them. Experiments were done, to choose the best possible parameters values for the effective collection of image dataset. Once the value was chosen, after performing a lot of experiments, these parameters (shown in Table I) fixed to specific values for each category.

Focus: Determining focus in an image is really important while dealing with complex images, especially if it is related to the biological field. These fungus images are hard to

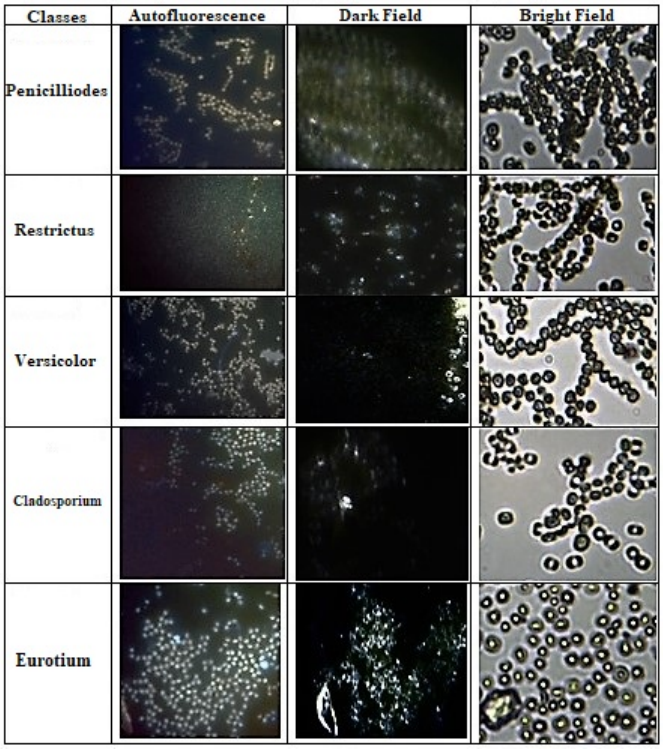


Fig. 3: Pictures of different kind of fungus spores in different light sources

Light Source	Gamma	Exposure	White Red Balance	Gain
Fluorescent	250	2.81	50	30
Dark Field	300	0.77	60	15
Bright Field	289	0.5	65	20

TABLE I: Parameters setting for different light sources

differentiate from each other if the field of view is not known, which certainly gives the effect of blurriness. This new change of blurriness was prevented by applying specific focus measure method. The focus measure is a technique to quantify the high-frequency detail in an image, and thereby giving a sharper focused image. It can be briefly understood by an equation 1.

$$\mathrm{c} = \frac{|O_d - F_d|}{O_d} \frac{f^2}{N(O_d - f)} \tag{1}$$

where c denotes the circle of confusion by having a blur circle with diameter c on the respective image plane. The distance from the out of focus object(spores) is denoted by O_d and distance from lens focus point to the lens is denoted by F_d. Aperture and f-number(f stop) is denoted by N and f respectively.

Generally, different objects appear blurred when captured randomly by a camera. Four different focus measure operators were applied and analyzed individually to measure the achievable results for obtaining sharp images.

1) Variance of Laplacian (LAP4): Measures the amount of edges present in the image.
2) Sum Modified Laplacian (SML): Computes local depth estimation at each image point.

Method	Image 1	Image 2	Image 3
Variance of Laplacian (LAP4)	68.43	65.45	66.70
Sum Modified Laplacian (SML)	3.35	3.23	3.65
Tenegrad (TENG)	588.66	604.09	594.85
MLOG	174	143	164

TABLE II: Focus Measure Methods

Fungus type	Image 1	Image 2	Image 3	Image 4	Image 5
Penicillioides	70.92	70.05	70.34	70.45	70.65
Restrictus	62.43	61.45	61.70	60.43	60.45
Versicolor	58.45	57.93	56.36	55.45	57.86
Cladosporium	68.43	65.45	66.70	64.43	64.00
Eurotium	72.59	71.95	72.10	71.46	72.35

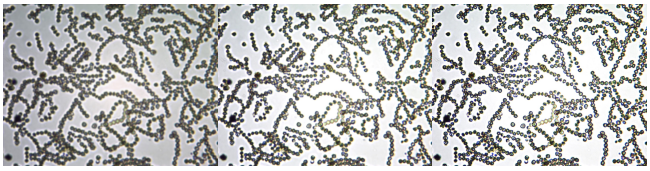
TABLE III: Variance of Laplacian (LAP4) values for each type of fungus

3) Tenengrad (TENG) focus measure operator: Based on the gradient of the image.
4) MLOG focus measure algorithm.

A brief outlook of these operators can be seen in Table II depicting corresponding values. Three test images of only one fungus type (e.g. Cladosporium) were taken in order to single out respective focus measure operator which was kept constant for taking all focused images of each fungus type in the dataset.

The variance of Laplacian (LAP4) was chosen in order to measure accurate focus of images because the blurred images are relatively indistinguishable from the focused images. Table III shows respective values of focus for each type of fungus for multiples images of the respective type. Six different images were captured for each fungus type, and only one with the highest value of LAP4 was selected, to avoid redundancy and time consumption. Depending upon these selected values, multiple images were taken in order to provide the constant and effective collection of the sharp and focused version of images.

Sharp images were acquired by replacing them with blurred images after applying above mentioned focus measure operator (LAP4). A quick glimpse is shown in Fig. 4 to depict visually that larger LAP4 value corresponds to sharper images.



(a) (LAP4 : 55.95) (b) (LAP4 : 57.04) (c) (LAP4 : 58.96)

Fig. 4: Blur and Sharp images of Cladosporium

D. Annotation of Dataset

Annotation is a time-consuming task but it is very important to accurately annotate fungus spore and dirt in the images. This task consumed plenty of time and required expertise. The annotators of fungus dataset were engineering students and they were trained to identify spores and their different types in the images, with the help of BMA labor Bochum. For the annotation purpose, a special Graphical User Inter(GUI) was also developed in MATLAB environment. This GUI can easily load, annotate and save dataset images.

GUI gives a stage to adequately point out boundaries around the fungus spores or dirt particles. It makes a record, which stores all the x-y coordinates of the marking. These coordinates help the annotator to take out the desired patch out of the image. GUI gives a lot of choices to annotate images by utilizing 10 unique sizes of drawing line and five distinct colors. The annotators can also redraw the annotation on their wish. In the long run, a mat file was generated, which has all the record of spores/ dirt particles coordinates present in the picture. The generated mat files containing pixel coordinates of the annotated fungus dirt and spore are also available in the dataset.

IV. Methodology

The convolutional neural network is a machine learning technique which is normally utilized for image, voice and data classification [34]. It comprises of multiple convolutional layers with pooling (subsampling) step and followed by one or more fully connected layers. Each layer receives an input, performs dot product with the weights of filters and then non-linearity is applied (optional). Filter weights and biases were learned with the help of minimization of the loss function. The whole CNN defines a single differentiable score function, which has input image on input end and class scores of belongingness on the other end. Scores were calculated with the help of loss function.

Our CNN architecture runs layer by layer as shown in Fig 5. These respective layers consist of neurons which have learnable weights (W) and biases (b). Which can be described by the equation

$$y = h_{W,b}(x) = f(W^T x) = f(\sum_{i=1}^{n} W_i x_i + b) \qquad (2)$$

where x is an input image with dimension $H\text{x}W\text{x}D$ and y is output with dimension $H'\text{x}W'\text{x}D'$. $h_{W,b}$ is the hypothesis and it is based on W and b. The parameters present in the first layer are denoted by W^1. The output of the first layer is used as an input for the second layer. And this process goes on till the last layer and the output of the last layer is y^l. This last layer is 6-dimensional vector because we have 6 classes(5 fungus classes and one dirt class) to classify. i^{th} element of this layer encloses the score of x^1(input image) belongs to i^{th} class. To calculate this score a layer was introduced at $L-1$ as a softmaxloss layer.

At the termination of CNN architecture, there is y^l, which is loss function/cost function. y' is the labeled output (ground truth) of the input image and it was utilized to measure any

discrepancy in CNN predicted output. The cost function can be defined as given below provided with the number of training examples with output labels as

$$\{(x^{(i)}, y^{(i)}) : i = 1, \ldots, m\} \tag{3}$$

where m is the number of training images.

$$J(W,b) = -\frac{1}{m}\sum_{i=1}^{m}\sum_{k=1}^{6}\Big(y_k^{(i)}\log(h_{W,b}(x^{(i)}))_k + (1-y_k^{(i)})\log(1-h_{W,b}(x^{(i)}))_k\Big) \tag{4}$$

where k attained value upto 6 because total number of classes which are going to be classified is 6 and only one of the two terms in the summation is non-zero for each training example. Moreover, this cost function $J(W,b)$ measures how well a given hypothesis $h_{W,b}$ fits the training data. It can learn to classify the training data by minimizing $J(W,b)$ to find the best choice of W and b. Where b can also be merged into W as one additional column. Partial derivative of this cost function $J(W)$ with respect to W_j can be written as:

$$\frac{\partial J(W)}{\partial W_j} = \sum_i x_j^{(i)}\Big(h_W(x^{(i)}) - y^{(i)}\Big) \tag{5}$$

writing in its vector form, the entire gradient can be expressed as :

$$\nabla_W J(W) = \sum_i x^{(i)}\Big(h_W(x^{(i)}) - y^{(i)}\Big) \tag{6}$$

The learning part of CNN was done through back propagation method, which is added as an additional layer. Back propagation [35] algorithm is commonly used for calculating the partial derivatives. One iteration of gradient descent updates the parameters W, b as follows:

$$W_{ij}^{(l)} = W_{ij}^{(l)} - \alpha\frac{\partial}{\partial W_{ij}^{(l)}}J(W,b) \tag{7}$$

$$b_i^{(l)} = b_i^{(l)} - \alpha\frac{\partial}{\partial b_i^{(l)}}J(W,b) \tag{8}$$

Given a training example (x, y), a complete forward pass was executed to compute all the activations throughout the network, including the output value of the hypothesis $h_{W,b}(x)$. Then, for each node i in layer l, an error term was computed. On the basis of this error, weights W and b were updated using equation 7 and equation 8.

A. CNN Architecture layers

In this section, we present the proposed convolutional neural network (CNN) architecture for the detection of fungus spores. Our architecture is summarized in Fig 5. It contained 11 layers in total where 3 were convolution, 3 were ReLU, 3 were Pooling and 2 were the fully connected layer. Below we describe the prominent characteristics of the different layers and how they were implemented.

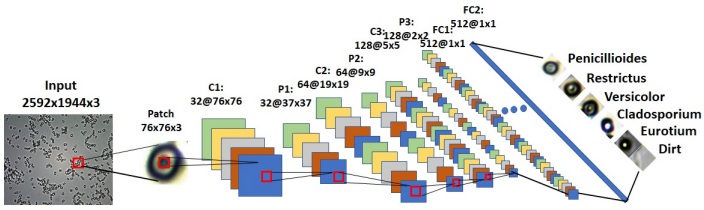


Fig. 5: CNN model for detection of different kind of fungus

1) Convolution Layer: Feature extraction is the main reason for applying convolution layer. It bears the main portion of the responsibility for automatic feature extraction. This feature can be extracted from an image by convolving a filter on the entire image. A new map comes out as shown in Fig 5. This new map contains the specific feature which is described in the applied filter. So, an image is convolved with multiple filters to extract multiple features out of the same image.

2) Pooling Layer: The resultant of the convolution layer is reduced in dimension as compared to input image as in Fig 5 but increased in depth. Depth is equal to the number of applied filters in the previous layer. The main purpose of the pooling layer is to reduce spatial size and number of parameters. This layer can be the input of the classifier directly but the answer will be still big and huge computational power will be utilized. Therefore, it is better to apply the pooling layer first to further reduce the dimensions. The pooling layer is also normally utilized to insert in between two convolution layers. There were two types of pooling were applied in the proposed architecture: max pooling and average pooling. In the case of the max pooling, the output of the depth slice of input is max value and in case of average pooling its average of all the values of the depth slice.

3) ReLU Layer: ReLU stands for Rectified Linear Units [21]. This is an activation function which is utilized in the proposed approach. The shape of this function similar to the ramp. It can be written as $I_{ijk} = max(0, x_{ijk})$. It has the advantage of less computation cost over other activation functions like sigmoid, tangent hyperbolic. In result of that, it greatly reduces the training time. So, the output of pooling layer exposed to non-linearity (ReLU) is shown in Fig 5.

4) Fully Connected Layer: After passing through all of the above mentioned layers, input fed into the fully connected layer. This layer is responsible for the high-level feature extraction. As shown in Fig 5, this layer has the most number of parameters and it has full connection with the activation of the previous layer.

5) Softmax Log Loss Layer: Right after the application of the fully connected layer, softmax log loss layer is introduced. This includes both softmax and logistic loss. Softmax layer is also a general type of fully connected layer. As described above that activation function used in fully connected is ReLU. In contrast to it, the softmax activation function, as shown in Eq. 9, is utilized in the softmax layer.

$$f(z)_i = \frac{\exp(z_j)}{\sum_{k=1}^{K}\exp(z_k)} \tag{9}$$

Layer	Type	Filter size	No. of Filters	Output size
0	Input	-	-	76x76x3
1	Conv	5x5x3	32	76x76x32
2	Max-Pool	2x2	32	37x37x32
3	ReLU	-	-	37x37x32
4	Conv	5x5x32	64	19x19x64
5	Max-Pool	2x2	64	9x9x64
6	ReLU	-	-	9x9x64
7	Conv	5x5x64	128	5x5x128
8	Max-Pool	2x2	128	2x2x128
9	ReLU	-	-	2x2x128
10	FC1	1x1x512	512	512
11	FC2	1x1x512	512	512
12	LogSoftmax	-	-	6

TABLE IV: Proposed CNN architecture for the fungus detection

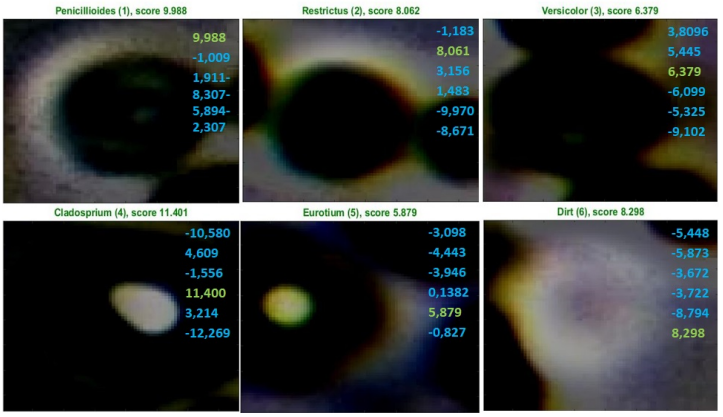


Fig. 6: Output of algorithm with score value

Which is finally responsible for classification task in convolution neural networks. Normally, it is incorporated where the major task is related to classification. It squashes the number of probabilities in a respective dimensional vector within the range of [0, 1]. In fact, it involves the generalization of logistic regression to handle multiple class problem. And to calculate the logistic loss of the function Eq. 10 was used.

$$L_i = -log(\frac{\exp(z_j)}{\sum_{k=1}^{K}\exp(z_k)}) \tag{10}$$

The architecture of designed CNN is shown in TABLE IV. It is designed to provide the score of belongingness of six classes as an output. It is empirically determined that eleven layers were best for the task of fungus detection. More layers can also be added but it will increase the computation cost and complexity of the approach and it will also not increase the accuracy. Number of layers, number of nodes in the layer and learning rate were hyper-parameters and were decided on the basis of best performance on our dataset. ReLU activation function is preferred for speedy training and to avoid vanishing gradient problem [21].

V. Results and Discussion

A. Dataset

As shown in Fig 1, we proposed a dataset of 5 different classes - Penicillioides, Restrictus, Versicolor, Cladosporium and Eurotium - and one extra class of dirt. This dataset comprises of 40,800 RGB images, which is further divided into two sets, training dataset and test dataset. The training set consisted of 30,000 fungus images with 5000 images per class. Test set consisted of 10,800 fungus spore images and dirt images, with 1800 images per class. Additionally, same amount of images were also collected using 2 other light sources. All three kind of images is shown in Fig. 4. Due to the image quality of other types of images, bright field images were preferred to apply CNN [36]. In total, this means several hundred thousand of pixels which are enough to train and test any machine learning technique for the purpose of classification and detection. Flipping and rotating can also be used as data augmentation technique if anyone requires more images for a sophisticated technique.

When more than 40 thousand images are available, deep learning technique can be easily applied to the task of image classification. As it showed state of the art results for the problem of image classification [37]. So, a CNN was trained using MATLAB and Geforce GTX 1060 6GB graphics processing unit (GPU) from NVIDIA was used. The network was assigned 0.01 as initial learning rate and was decreased by a factor of 10 when the validation error was approximately constant. CNN was trained for the 20 epochs. The whole dataset was divided into 5 folds because 5 fold validation scheme was used. This signifies that 20% of the data was used for the purpose of validating the results and the fine tuning of classifier hyperparameters.

B. Algorithmic Analysis

The performance of the proposed method was evaluated on the 10,800 images of test dataset. Which has 1800 images of each class. The output of the proposed architecture on fungus dataset is shown in Fig. 6. This figure contains the output of the testing and the score values. Furthermore, the score value of other classes is also shown on each output. This score means that our algorithm is how much sure about the input is of that class. For example, Penicillioides score value is 9.988, which is very less in number but if we compare it with other classes scores(-1.01, 1.91, -8.31, -5.89, -2.31) then it is much higher. Moreover, there is a great difference between 9.988 and 1.911, which is the second best option. This means that proposed system not only accurately detect correct class but also with a good margin in the score. Scores of all 6 examples are shown in Fig. 7. Furthermore, chart is also provided to compare it with score values of other fungus classes.

It is also very interesting to see the false classified results as shown in Fig. 8. Two example cases are displayed in the figure. Versicolor was the falsely classified as Restrictus with the score of 4.507 but it is also observed from the score values that Restrictus has the second best score of 2.71. Which is 1.8 short of the best score. Similarly in the second case, Restrictus was falsely classified as Versicolor with the 9.468 score value and the second best score is 8.83 which is just short of the best score. And this score belongs to

	1	2	3	4	5	6
Penicillioides	9.988	-1.183	3.8096	-10.58	-3.098	-5.448
Restrictus	-1.009	8.061	5.445	4.609	-4.443	-5.873
Versicolor	1.911	3.156	6.379	-1.556	-3.946	-3.672
Cladosprium	8.307	1.483	-6.099	11.4	0.1382	-3.722
Eurotium	5.894	-9.97	-5.325	3.214	5.879	-8.794
Dirt	2.307	-8.671	-9.102	-12.269	-0.827	8.298

Fig. 7: Score values of some test examples

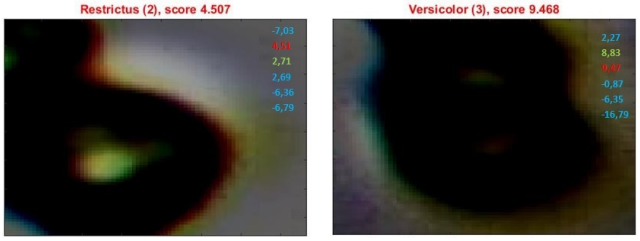


Fig. 8: False classified test images

the correct class (Restrictus). The remaining scores of other classes are way behind these score values.

During this complete process of training and testing, various observations were noticed about some parameters. Which were also helpful in enhancing the model performance by varying their values. These parameters are briefly described further with respect to the results.

1) Learning Rate: The most important hyperparameter for improving the accuracy of the model is learning rate. Different strategies were applied in recent years to improve the accuracy with help of the learning rate. Most of the models accept variable learning rates in each layer according to the behavior shown by training processes and some retain the constant values throughout the training. It has been observed that keeping the constant value throughout the training process achieve the acceptable accuracy as shown in Fig 9. But the learning rate used for designed architecture was not constant, it was decayed exponentially as the training progressed. The starting value of 0.01 was selected after many experiments. As shown in Fig 9

2) Weight Initialization: Currently there exists no rule of thumb to initialize the weights of the hidden layer filters through the particular format. Different experiments were performed and the accuracy was monitored. [38] and [39] recommended scaling by the inverse of the square root of the number of inputs. Whereas, [40] suggested using both

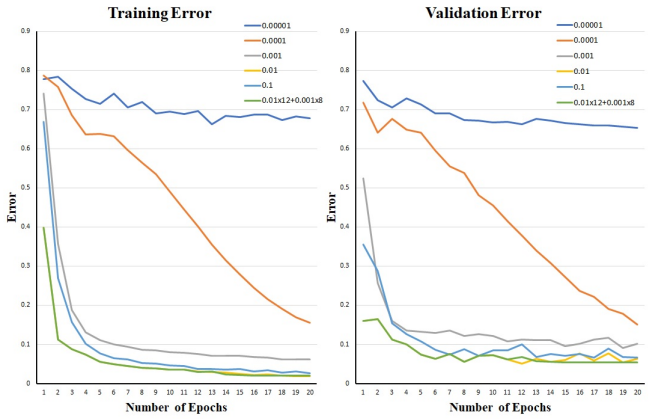


Fig. 9: Training and validation error at different learning rate

numbers of inputs and number of hidden units. Four weight initialization techniques were found in the recent literature and all of them were tried and found that D performs best on fungus detection dataset. The reason behind in selecting the interval D, is to initialize the weights as smaller enough around the origin resulting that activation function can operate in a linear regime, where the gradients have largest values.

$$\mathrm{A} = (0, 0) \tag{11}$$

$$\mathrm{B} = [-4r, +4r] \tag{12}$$

$$\mathrm{C} = [-2r, +2r] \tag{13}$$

$$\mathrm{D} = [-r, r] \tag{14}$$

where r is

$$r = [\frac{\sqrt{6}}{fan_{in} + fan_{out}}, \frac{\sqrt{6}}{fan_{in} + fan_{out}}] \tag{15}$$

Where fan_{in} denotes the number of inputs and Fan_{out} denotes the number of hidden units. The reason behind in selecting the interval D, is to initialize the weights as smaller enough around the origin resulting that activation function can operate in a linear regime, where the gradients have largest values.

3) Number of hidden layers: As far as number of hidden layers are concerned, these are totally dependent on image dataset. This parameter can handle a lot of other parameters even if the input contribution of the dataset is complicated, thereby acts accordingly in order to stabilize the model. More capacity of hidden layers is required to model more complex input structure. This hyperparameter is closely related to weight initialization, as weights are initialized in some respective layers, which control the number of inputs and number of hidden units, finally resulting in the reduction of parameters which save computational effort. It has been observed that by increasing number of hidden layers, the accuracy of the model was increased. Different measurements were performed with the wise selection of number of hidden

	Zhang et al. [41]	Proposed approach
Application	Fungi of Leucorrhea images	Fungi of food, fungi of Archives, Fungi in the environment
Type of fungi	Identifying only one type	Successfully classifying five different type of fungus
Collection of images	Manual	Through special optical sensor system
Type of images	Only one type of images was obtained	Three types of images were obtained
Presented Dataset	No	Yes
Segmentation technique	Threshold based Segmentation	CNN
Feature Extraction technique	Histogram of Oriented Gradients	CNN
Classification technique	Support Vector Machines	CNN
Dimensionality reduction required	Yes	No
CNN architecture	7 layers	11 layers
Stride	Fixed value 1	Value 1 and value 2 used as per requirement
Pooling layer type	Average pooling	Max pooling
Activation function	Sigmoid function	ReLU function
Loss function	Cross entropy loss	Softmax log loss

TABLE V: Comparison of proposed approach with Zhang et al.

layers. We have used 11 number of hidden layers in total for our model.

4) Mini Batch Size: Mini batch size also affect the final accuracy of the CNN. Negative trend is observed during our experiments when different mini batch sizes experimented with constant learning rate. It is recommended that learning rate and mini batch size values should be selected in combination.

5) Number of filters and size of filters: Besides above explained parameters, number of filter and size of filters also define the training model accuracy. Different filter sizes experimented in convolution layers and it has been observed that increasing the number of filters with smaller size improves the accuracy.

C. Comparison with other fungus detection CNN technique

Recently Zhang et al. [41] presented an approach for fungi identification in Leucorrhea images. They presented approach for the identification of one type of fungi. It is interesting to compare our approach with Zhang et al. Table V enumerates the differences between both approaches.

Zhang et al. manually collected microscopic images of fungal hyphae, whereas in proposed approach fungus images are collected through completely automated optical sensor system. They collected only bright field microscopic images on the other hand fungus detection system collected, bright field, dark field and autofluorescence microscopic images. Moreover, in this paper complete dataset is also presented.

The second major difference is the variable approaches. They used traditional computer vision approach which consists of pre-processing, segmentation, feature extraction and classification. Threshold-based segmentation was used which is followed by CNN. Then features were extracted through histogram of oriented gradients and at last SVM was used for classification. Which increased the computational complexity of their approach and make it less efficient as compared to our approach, where only CNN was used for all these steps (preprocessing, segmentation, feature extraction, classification). Moreover, there are lot of differences in the CNN architecture which are also summarizes in Table V.

VI. Conclusion

In this paper, we have introduced a novel fungus dataset for the detection of fungus, that poses severe health problems to human life and huge financial loss to the food and agriculture industry. Our dataset contains 40,800 images of 6 classes. It is further divided into 10,800 test images and 30,000 train images. These images were captured with the help of complete automatic fungus detection system. We plan to add more images and their annotations to this dataset. We have also proposed a technique for fungus detection using CNN that has achieved 94.8% accuracy with 5 fold validation. Which not only outclasses with 6.8% improvement in detection but also it has the ability to differentiate different kinds of fungus. Through these results, it can be concluded that making of fungus dataset and the proposed CNN technique is an important step towards an accurate, reliable and early detection system for fungus. Furthermore, these findings motivate us to apply the proposed approach to indoor air monitoring and other microbial in the environment.

In our future research, we will explore regional convolutional neural network and transfer learning for the fungus detection. Earlier studies show that success of deep learning techniques are on larger dataset but transfer learning can be helpful for the small dataset. We are also modifying our current fungus detection system, which will also improve picture quality. For these pictures, we will apply transfer learning by removing the last few layers and then include new layers in the architecture for the fine tuning, majorly based on the proposed model. Moreover, a trained network will be implemented on other types of images(air born particles, pollens etc) after fine tuning.

References

[1] J. I. Pitt, A. D. Hocking *et al.*, *Fungi and food spoilage*. Springer, 2009, vol. 519.

[2] J. Gustavsson, C. Cederberg, U. Sonesson, R. van Otterdijk, and A. Meybeck, "Global food losses and food waste: extent, causes and prevention. fao, rome," 2011.

[3] P. Papireddy Vinayaka, S. van den Driesche, R. Blank, M. W. Tahir, M. Frodl, W. Lang, and M. J. Vellekoop, "An impedance-based mold sensor with on-chip optical reference," *Sensors*, vol. 16, no. 10, p. 1603, 2016.

[4] R. Blank, P. Vinayaka, M. Tahir, J. Yong, M. Vellekoop, and W. Lang, "Development of a fungal risk monitor for the next generation of intelligent containersaper," in *2016 IEEE SENSORS*. IEEE, 2016, pp. 1–3.

[5] P. P. Vinayaka, S. van den Driesche, R. Blank, A. Chakraborty, R. Amin, M. Tahir, N. Zaidi, M. Frodl, W. Lang, and M. Vellekoop, "Membrane-sealed bioreactor for on-site autonomous detection of fungi spore contamination in archives," *Procedia Engineering*, vol. 168, pp. 529–532, 2016.

[6] P. P. Vinayaka, R. Blank, M. Frodl, M. W. Tahir, N. A. Zaidi, M. Oellers, S. van den Driesche, W. Lang, and M. J. Vellekoop, "A replaceable bioreactor array for monitoring mold spore contamination in archives," in *MikroSystemTechnik 2017; Congress; Proceedings of*. VDE, 2017, pp. 641–644.

[7] Y. Wang, Y. Yin, and C. Zhang, "Selective cultivation and rapid detection of staphylococcus aureus by computer vision," *Journal of food science*, vol. 79, no. 3, 2014.

[8] J. Blasco, N. Aleixos, and E. Moltó, "Computer vision detection of peel defects in citrus by means of a region oriented segmentation algorithm," *Journal of Food Engineering*, vol. 81, no. 3, pp. 535–543, 2007.

[9] K. Mertens, B. De Ketelaere, B. Kamers, F. Bamelis, B. Kemps, E. Verhoelst, J. De Baerdemaeker, and E. Decuypere, "Dirt detection on brown eggs by means of color computer vision," *Poultry science*, vol. 84, no. 10, pp. 1653–1659, 2005.

[10] M. W. Tahir, N. Zaidi, R. Blank, P. Vinayaka, M. Vellekoop, and W. Lang, "Detection of fungus through an optical sensor system using the histogram of oriented gradients," in *2016 IEEE SENSORS*. IEEE, 2016, pp. 1–3.

[11] S. Gunasekaran, "Computer vision technology for food quality assurance," *Trends in Food Science & Technology*, vol. 7, no. 8, pp. 245–256, 1996.

[12] K. K. Reddy and M. Shah, "Recognizing 50 human action categories of web videos," *Machine Vision and Applications*, vol. 24, no. 5, pp. 971–981, 2013.

[13] K. Soomro, A. R. Zamir, and M. Shah, "Ucf101: A dataset of 101 human actions classes from videos in the wild," *arXiv preprint arXiv:1212.0402*, 2012.

[14] A. Krizhevsky and G. Hinton, "Convolutional deep belief networks on cifar-10," *Unpublished manuscript*, vol. 40, p. 7, 2010.

[15] A. Krizhevsky, , and G. Hinton, "Learning multiple layers of features from tiny images," 2009.

[16] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *IEEE Conference on Computer Vision and Pattern Recognition, 2009. CVPR 2009*. IEEE, 2009, pp. 248–255.

[17] K. Sirinukunwattana, S. E. A. Raza, Y.-W. Tsang, D. R. Snead, I. A. Cree, and N. M. Rajpoot, "Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1196–1206, 2016.

[18] S. Wienert, D. Heim, K. Saeger, A. Stenzinger, M. Beil, P. Hufnagl, M. Dietel, C. Denkert, and F. Klauschen, "Detection and segmentation of cell nuclei in virtual microscopy images: a minimum-model approach," *Scientific reports*, vol. 2, p. 503, 2012.

[19] E. D. Gelasca, B. Obara, D. Fedorov, K. Kvilekval, and B. Manjunath, "A biosegmentation benchmark for evaluation of bioimage analysis methods," *BMC bioinformatics*, vol. 10, no. 1, p. 368, 2009.

[20] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[21] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[22] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *European Conference on Computer Vision*. Springer, 2014, pp. 818–833.

[23] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9.

[24] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *arXiv preprint arXiv:1512.03385*, 2015.

[26] B. Dong, L. Shao, M. Da Costa, O. Bandmann, and A. F. Frangi, "Deep learning for automatic cell detection in wide-field microscopy zebrafish images," in *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2015, pp. 772–776.

[27] T. Chen and C. Chefdhotel, "Deep learning based automatic immune cell detection for immunohistochemistry images," in *International Workshop on Machine Learning in Medical Imaging*. Springer, 2014, pp. 17–24.

[28] C. D. Malon and E. Cosatto, "Classification of mitotic figures with convolutional neural networks and seeded blob features," *Journal of pathology informatics*, vol. 4, 2013.

[29] U. H. GmbH, "Mbass30: Air sampling unit," https://www.holbach.biz/upload/manual_mbass30_v2.pdf, accessed: 2017-06-15.

[30] R. Blank, P. Vinayaka, M. Tahir, M. Vellekoop, and W. Lang, "Optical sensor system for the detection of mold," in *IEEE Sensors*, 2015.

[31] M. W. Tahir, N. A. Zaidi, R. Blank, P. P. Vinayaka, M. J. Vellekoop, and W. Lang, "Fungus detection through optical sensor system using two different kinds of feature vectors for the classification," *IEEE Sensors Journal*, vol. 17, no. 16, pp. 5341–5349, 2017.

[32] M. W. Tahir, N. A. Zaidi, R. Blank, P. Vinayaka, M. Vellekoop, and W. Lang, "An efficient and simple embedded system of fungus detection system," in *2017 International Multi-topic Conference (INMIC)*. IEEE, 2017, pp. 1–4.

[33] M. W. Tahir, N. Zaidi, R. Blank, P. P. Vinayaka, and W. Lang, "Fungus detection system," in *2016 IEEE International Conference on Autonomic Computing (ICAC)*. IEEE, 2016, pp. 227–228.

[34] T. Liu, S. Fang, Y. Zhao, P. Wang, and J. Zhang, "Implementation of training convolutional neural networks," *arXiv preprint arXiv:1506.01195*, 2015.

[35] S. Dewan and S. Chakravarthy, "A system for offline character recognition using auto-encoder networks," in *International Conference on Neural Information Processing*. Springer, 2012, pp. 91–99.

[36] M. W. Tahir, N. A. Zaidi, R. Blank, P. P. Vinayaka, M. J. Vellekoop, and W. Lang, "Comparison of pre-processing on different kind of images produced by optical sensor system," in *2nd International Conference on Sensors Engineering and Electronics Instrumental Advances (SEIA' 2016)*. IFSA, 2016, pp. 41–43.

[37] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[38] Y. Bengio, "Practical recommendations for gradient-based training of deep architectures," in *Neural networks: Tricks of the trade*. Springer, 2012, pp. 437–478.

[39] P. Simard, Y. LeCun, J. Denker, and B. Victorri, "Transformation invariance in pattern recognition tangent distance and tangent propagation," *Neural networks: tricks of the trade*, pp. 549–550, 1998.

[40] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, 2010, pp. 249–256.

[41] J. Zhang, S. Lu, X. Wang, X. Du, G. Ni, J. Liu, L. Liu, and Y. Liu, "Automatic identification of fungi in microscopic leucorrhea images," *JOSA A*, vol. 34, no. 9, pp. 1484–1489, 2017.

Muhammad Waseem Tahir received the B.Sc. and MS degree in electrical engineering from Riphah International University, Islamabad, Pakistan, in 2009 and 2013, respectively. He also worked as lecturer in the Riphah International University for 5 years. Then he came to Germany for his PhD. Currently, he is pursuing the Ph.D. degree with the Institute for Microsensors, -Actuators and -Systems (IMSAS), University of Bremen, Germany. He also worked as guest researcher in Center of Research in Computer Vision (CRCV), University of Central Florida, USA. He has co-authored many research articles in the field of robotics, medical imaging, computer vision, bio sensors and sensor systems. In addition to above topics his current interests include machine learning.