CHEST DISEASE DETECTION USING CNN AI ALGORITHM

Borapureddi Mohan Krishna[1], Palancha Uday[2], Reddymalla Jashua[3], Sravanthi B[4]
[1,2,3] UG Scholar, Dept. of AI & ML, St. Martin's Engineering College, Secunderabad, Telangana, India, 500100
[4] Assistant Professor, Dept. of AI & ML, St. Martin's Engineering College, Secunderabad, Telangana, India, 500100
borapureddimohan@gmail.com

Abstract:

Pneumonia is a life-threatening infectious disease affecting one or both lungs in humans commonly caused by bacteria called Streptococcus pneumoniae. One in three deaths in India is caused due to pneumonia as reported by World Health Organization (WHO). Chest X-Rays which are used to diagnose pneumonia need expert radiotherapists for evaluation. Thus, developing an automatic system for detecting pneumonia would be beneficial for treating the disease without any delay particularly in remote areas. Due to the success of deep learning algorithms in analyzing medical images, Convolutional Neural Networks (CNNs) have gained much attention for disease classification. In addition, features learned by pre-trained CNN models on large-scale datasets are much useful in image classification tasks. In this work, we appraise the functionality of pre-trained CNN models utilized as feature-extractors followed by different classifiers for the classification of abnormal and normal chest X-Rays. We analytically determine the optimal CNN model for the purpose. Statistical results obtained demonstrates that pretrained CNN models employed along with supervised classifier algorithms can be very beneficial in analyzing chest X-ray images, specifically to detect Pneumonia.the study explores various pre-trained CNN architectures such as VGG16, ResNet50, and InceptionV3 to identify the most effective model for pneumonia detection

Keywords: Federated Learning (FL), Massive open online course (MOOCs), Small Private Online Course(SPOCs), IoT based industries, retention assessment.

1.INTRODUCTION

X-ray radiographies are an affordable and non-invasive method of examining different organs of the body. Recognized as a valuable diagnosis tool for many disorders and abnormalities, X-rays can also be used to monitor diseases during treatment. Around 3.6 billion X-ray images are taken every year worldwide. This number includes over 150 million chest X-ray radiographies (CXR) performed in the United States only. The World Health Organization (WHO) stated that CXRs are the most commonly performed clinical imaging technique worldwide. CXRs are grayscale images generally produced by projecting X-rays onto the human body positioned against a metallic plate. Although CXRs play a crucial role in the diagnosis of thoracic disease, visual inspection by radiologists remains complex and error-prone. Previous studies have shown that the risk of misdiagnosis increases with the amount of time it takes for a radiologist to interpret CXR images. In addition, even experienced radiologists were at greater risk of misdiagnosis because of hidden lesions and symptoms in soft tissue and bones. The WHO reports that many chest diseases can be life threatening and lead to the death of millions of people if not accurately and timely treated. Some chest diseases are of high mortality rates such as tuberculosis that kills around 1.4 million people annually, pneumonia that kills 9 million children under the age of 5 years being the world's leading killer disease, and COVID-19 which caused the death of over 6 million people all over the world as of November 2022.

In recent years, the detection of chest diseases using Convolutional Neural Networks (CNNs) has become a groundbreaking approach in the field of medical diagnostics. CNNs, a subset of deep learning algorithms, excel at processing and analyzing visual data, making them particularly suitable for interpreting medical images such as chest X-rays and CT scans. These networks are designed to automatically learn and extract features from input images through multiple layers of convolutional filters, which can identify patterns and anomalies indicative of various chest diseases, including pneumonia, tuberculosis, and lung cancer. By training CNNs on large datasets of labeled medical images, they can achieve remarkable accuracy in disease detection, often surpassing traditional methods and even human radiologists in some cases. This advancement not only enhances diagnostic precision but also accelerates the process, enabling timely and effective treatment for patients. The integration of CNNs into medical imaging workflows is poised to revolutionize healthcare, providing powerful tools for early detection and intervention, ultimately improving patient outcomes and reducing the burden on healthcare systems

.

2. LITERATURE SURVEY

[1] "Machine learning approach for automated screening of malaria parasite using light microscopic images

The aim of this paper is to address the development of computer assisted malaria parasite characterization and classification using machine learning approach based on light microscopic images of peripheral blood smears. In doing this, microscopic image acquisition from stained slides, illumination correction and noise reduction, erythrocyte segmentation, feature extraction, feature selection and finally classification of different stages of malaria (Plasmodium vivax and Plasmodium falciparum) have been investigated. The erythrocytes are segmented using marker controlled watershed transformation and subsequently total ninety six features describing shape-size and texture of erythrocytes are extracted in respect to the parasitemia infected versus non-infected cells. Ninety four features are found to be statistically significant in discriminating six classes. Here a feature selection-cum-classification scheme has been devised by combining F-statistic, statistical learning techniques i.e., Bayesian learning and support vector machine (SVM) in order to provide the higher classification accuracy using best set of discriminating features. Results show that Bayesian approach provides the highest accuracy i.e., 84% for malaria classification by selecting 19 most significant features while SVM provides highest accuracy i.e., 83.5% with 9 most significant features. Finally, the performance of these two classifiers under feature selection framework has been compared toward malaria parasite classification.

[2] "Image analysis and machine learning for detecting malaria"

Malaria remains a major burden on global health, with roughly 200 million cases worldwide and more than 400,000 deaths per year. Besides biomedical research and political efforts, modern information technology is playing a key role in many attempts at fighting the disease. One of the barriers toward a successful mortality reduction has been inadequate malaria diagnosis in particular. To improve diagnosis, image analysis software and machine learning methods have been used to quantify parasitemia in microscopic blood slides. This article gives an overview of these techniques and discusses the current developments in image analysis and machine learning for microscopic malaria diagnosis. We organize the different approaches published in the literature according to the techniques used for imaging, image preprocessing, parasite detection and cell segmentation, feature computation, and automatic cell classification. Readers will find the different techniques listed in tables, with the relevant articles cited next to them, for both thin and thick blood smear images. We also discussed the latest developments in sections devoted to deep learning and smartphone technology for future malaria diagnosis

[3] "Automated image processing method for the diagnosis and classification of malaria on thin blood smears."

Malaria is a serious global health problem, and rapid, accurate diagnosis is required to control the disease. An image processing algorithm to automate the diagnosis of malaria on thin blood smears is developed. The image classification system is designed to positively identify malaria parasites present in thin blood smears, and differentiate the species of malaria. Images are acquired using a charge-coupled device camera connected to a light microscope. Morphological and novel threshold selection techniques are used to identify erythrocytes (red blood cells) and possible parasites present on microscopic slides. Image features based on colour, texture and the geometry of the cells and parasites are generated, as well as features that make use of a priori knowledge of the classification problem and mimic features used by human technicians. A two-stage tree classifier using backpropogation feedforward neural networks distinguishes between true and false positives, and then diagnoses the species (Plasmodium falciparum, P. vivax, P. ovale or P. malariae) of the infection. Malaria samples obtained from the Department of Clinical Microbiology and Infectious Diseases at the University of the Witwatersrand Medical School are used for training and testing of the system. Infected erythrocytes are positively identified with a sensitivity of 85% and a positive predictive value (PPV) of 81%, which makes the method highly sensitive at diagnosing a complete sample provided many views are analysed. Species were correctly determined for 11 out of 15 samples.

[4] " CNN features off-the-shelf: an astounding baseline for recognition."

Recent results indicate that the generic descriptors extracted from the convolutional neural networks are very powerful. This paper adds to the mounting evidence that this is indeed the case. We report on a series of experiments conducted for different recognition tasks using the publicly available code and model of the OverFeat network which was trained to perform object classification on ILSVRC13. We use features extracted from the OverFeat network as a generic image representation to tackle the diverse range of recognition tasks of object image classification, scene recognition, fine grained recognition, attribute detection and image retrieval applied to a diverse set of datasets. We selected these tasks and datasets as they gradually move further away from the original task and data the OverFeat network was trained to solve. Astonishingly, we report consistent superior results compared to the highly tuned state-of-the-art systems in all the visual classification tasks on various datasets. For instance retrieval it consistently outperforms low memory footprint methods except for sculptures dataset. The results are achieved using a linear SVM classifier (or L2 distance in case of retrieval) applied to a feature representation of size 4096 extracted from a layer in the net. The representations are further modified using simple augmentation techniques e.g. jittering. The results strongly suggest that features obtained from deep learning with convolutional nets should be the primary candidate in most visual recognition tasks.

[5] "Imagenet classification with deep convolutional neural networks"

We trained a large, deep convolutional neural network to classify the 1.3 million high-resolution images in the LSVRC-2010 ImageNet training set into the 1000 different classes. On the test data, we achieved top-1 and top-5 error rates of 39.7\% and 18.9\% which is considerably better than the previous state-of-the-art results. The neural network, which has 60 million parameters and 500,000 neurons, consists of five convolutional layers, some of which are followed by max-pooling layers, and two globally connected layers with a final 1000-way softmax. To make training faster, we used non-saturating neurons and a very efficient GPU implementation of convolutional nets. To reduce overfitting in the globally connected layers we employed a new regularization method that proved to be very effective. Here a feature selection-cum-classification scheme has been devised by combining F-statistic, statistical learning techniques i.e., Bayesian learning and support vector machine (SVM) in order to provide the higher classification accuracy using best set of discriminating features. Results show that Bayesian approach provides the highest accuracy i.e., 84% for malaria classification by selecting 19 most significant features while SVM provides highest accuracy i.e., 83.5% with 9 most significant features. Finally, the performance of these two classifiers under feature selection framework has been compared toward malaria parasite classification.

[6]" Very deep convolutional networks for large-scale image recognition"

In this work we investigate the effect of the convolutional network depth on its accuracy in the large-scale image recognition setting. Our main contribution is a thorough evaluation of networks of increasing depth, which shows that a significant improvement on the prior-art configurations can be achieved by pushing the depth to 16-19 weight layers. These findings were the basis of our ImageNet Challenge 2014 submission, where our team secured the first and the second places in the localisation and classification tracks respectively.Recent results indicate that the generic descriptors extracted from the convolutional neural networks are very powerful. This paper adds to the mounting evidence that this is indeed the case. We report on a series of experiments conducted for different recognition tasks using the publicly available code and model of the OverFeat network which was trained to perform object classification on ILSVRC13. We use features extracted from the OverFeat network as a generic image representation to tackle the diverse range of recognition tasks of object image classification, scene recognition, fine grained recognition, attribute detection and image retrieval applied to a diverse set of datasets. We selected these tasks and datasets as they gradually move further away from the original task and data the OverFeat network was trained to solve. Astonishingly, we report consistent superior results compared to the highly tuned state-of-the-art systems in all the visual classification tasks on various datasets

[7] "Xception: deep learning with separable convolutions"

Malaria is a blood disease caused by the Plasmodium parasites transmitted through the bite of female Anopheles mosquito. Microscopists commonly examine thick and thin blood smears to diagnose disease and compute parasitemia. However, their accuracy depends on smear quality and expertise in classifying and counting parasitized and uninfected cells. Such an examination could be arduous for large-scale diagnoses resulting in poor quality. State-of-the-art image-analysis based computeraided diagnosis (CADx) methods using machine learning (ML) techniques, applied to microscopic images of the smears using hand-engineered features demand expertise in analyzing morphological, textural, and positional variations of the region of interest (ROI). In contrast, Convolutional Neural Networks (CNN), a class of deep learning (DL) models promise highly scalable and superior results with end-to-end feature extraction and classification. Automated malaria screening using DL techniques could, therefore, serve as an effective diagnostic aid. In this study, we evaluate the performance of pre-trained CNN based DL models as feature extractors toward classifying parasitized and uninfected cells to aid in improved disease screening. We experimentally determine the optimal model layers for feature extraction from the underlying data. Statistical validation of the results demonstrates the use of pre-trained CNNs as a promising tool for feature extraction for this purpose.

[8] “Deep residual learning for image recognition ”

Deeper neural networks are more difficult to train. We present a residual learning framework to ease the training of networks that are substantially deeper than those used previously. We explicitly reformulate the layers as learning residual functions with reference to the layer inputs, instead of learning unreferenced functions. We provide comprehensive empirical evidence showing that these residual networks are easier to optimize, and can gain accuracy from considerably increased depth. On the ImageNet dataset we evaluate residual nets with a depth of up to 152 layers - 8× deeper than VGG nets [40] but still having lower complexity. An ensemble of these residual nets achieves 3.57% error on the ImageNet test set. This result won the 1st place on the ILSVRC 2015 classification task. We also present analysis on CIFAR-10 with 100 and 1000 layers. The depth of representations is of central importance for many visual recognition tasks. Solely due to our extremely deep representations, we obtain a 28% relative improvement on the COCO object detection dataset. Deep residual nets are foundations of our submissions to ILSVRC & COCO 2015 competitions1, where we also won the 1st places on the tasks of ImageNet detection, ImageNet localization, COCO detection, and COCO segmentation.

[9] " Densely Connected Convolutional Networks"

Recent work has shown that convolutional networks can be substantially deeper, more accurate, and efficient to train if they contain shorter connections between layers close to the input and those close to the output. In this paper, we embrace this observation and introduce the Dense Convolutional Network (DenseNet), which connects each layer to every other layer in a feed-forward fashion. Whereas traditional convolutional networks with L layers have L connections - one between each layer and its subsequent layer - our network has L(L+1)/2 direct connections. For each layer, the feature-maps of all preceding layers are used as inputs, and its own feature-maps are used as inputs into all subsequent layers. DenseNets have several compelling advantages: they alleviate the vanishing-gradient problem, strengthen feature propagation, encourage feature reuse, and substantially reduce the number of parameters. We evaluate our proposed architecture on four highly competitive object recognition benchmark tasks (CIFAR-10, CIFAR-100, SVHN, and ImageNet). DenseNets obtain significant improvements over the state-of-the-art on most of them, whilst requiring less memory and computation to achieve high performance. Code and models are available at

[10] “Classification using deep learning neural networks for brain tumors”

Pneumonia is a life-threatening infectious disease affecting one or both lungs in humans commonly caused by bacteria called Streptococcus pneumonia. The present study aimed to examine the risk factors for death due to pneumonia in young children. One or more in three deaths in Asia is caused due to pneumonia as reported by World Health Organization (WHO). Chest X-Rays which are used to diagnose pneumonia need expert radiotherapists for evaluation. Thus, developing an automatic system for detecting pneumonia would be beneficial and it can save lots of peoples life and help stopping and curing and controll for treating the disease without any delay particularly in remote areas. Due to the success of deep learning algorithms in analyzing medical images, Convolutional Neural Networks (CNNs) have gained much attention for disease classification. In addition, features learned by pre-trained CNN models on large-scale datasets are much useful in image classification tasks. In this work, we appraise the functionality of pre-trained CNN models utilized as feature-extractors followed by different classifiers for the classification of abnormal and normal chest X-Rays. We analytically determine the optimal CNN model for the purpose. Statistical results obtained demonstrates that pretrained CNN models employed along with supervised classifier algorithms can be very beneficial in analyzing chest X-ray images, specifically to detect Pneumonia.

3. PROPOSED METHODOLOGY

In this project as per your request we have employed CNN and traditional machine learning algorithms like Random Forest and Decision tree to detect various viral chest disease infections. Each algorithm performance is evaluated in terms of precision, recall, accuracy, confusion matrix and FSCORE. Among all algorithms CNN is giving best detection accuracy. The primary goal of using Convolutional Neural Network in most of the image classification tasks is to reduce the computational complexity of the model which is likely to increase if the input are images . The original 3-channel images were resized from 1024×1024 into 224×224 pixels to reduce the heavy computation and for faster processing. All of the further techniques has been applied over these downsized images.

Advantages:

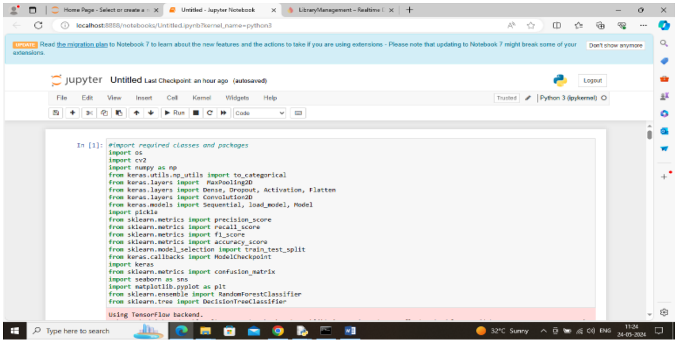
1. High Accuracy
2. Takes less time
3. Early Identifications of struggling students.
4. Improved retention rates.

Applications:

1. **Disease Diagnosis**: CNNs are used to detect conditions such as pneumonia, tuberculosis, lung cancer, and COVID-19 by analyzing chest X-rays or CT scans2.
2. **Early Detection**: They enable early diagnosis of diseases, improving treatment outcomes and reducing mortality rates.
3. **Automated Screening**: CNNs assist in large-scale screening programs, especially in areas with limited access to radiologists.

4. EXPERIMENTAL ANALYSIS

In above screen click on ‘Upload OULAD Online Student Dataset’ button to upload dataset and get below page



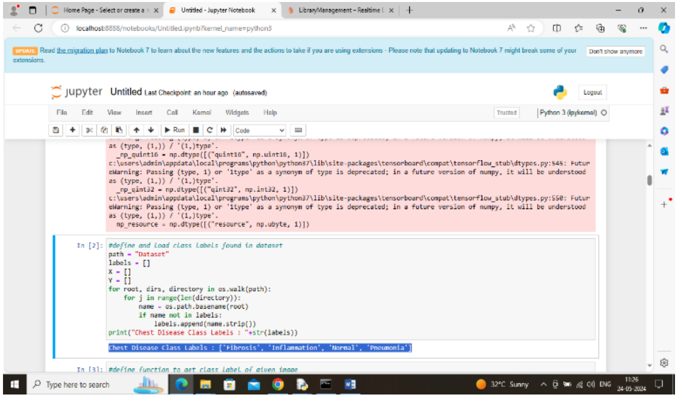
Read the migration plan to Notebook 7 to learn about the new features and the actions to take if you are using extensions - Please note that updating to Notebook 7 might break some of your extensions.

```
In [1]: #import required classes and packages
import os
import cv2
import numpy as np
from keras.utils.np_utils import to_categorical
from keras.layers import  MaxPooling2D
from keras.layers import Dense, Dropout, Activation, Flatten
from keras.layers import Convolution2D
from keras.models import Sequential, load_model, Model
import pickle
from sklearn.metrics import precision_score
from sklearn.metrics import recall_score
from sklearn.metrics import f1_score
from sklearn.metrics import accuracy_score
from sklearn.model_selection import train_test_split
from keras.callbacks import ModelCheckpoint
import keras
from sklearn.metrics import confusion_matrix
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.ensemble import RandomForestClassifier
from sklearn.tree import DecisionTreeClassifier

Using TensorFlow backend.
```

Fig. 1. Dataset Folder

In above screen defining function to identify class labels available in dataset



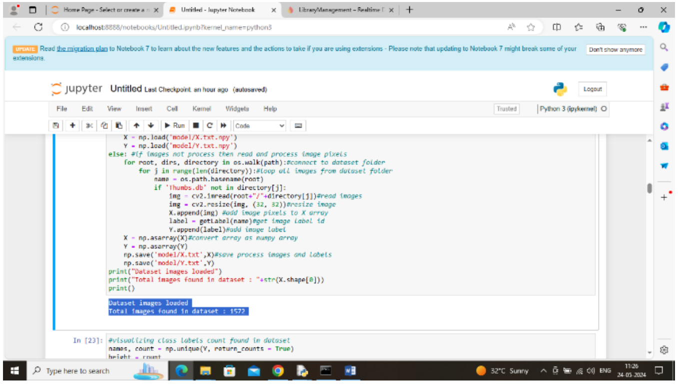
```
as (type, (1,)) / '(1,)type'.
  _np_quint16 = np.dtype([("quint16", np.uint16, 1)])
c:\users\admin\appdata\local\programs\python\python37\lib\site-packages\tensorboard\compat\tensorflow_stub\dtypes.py:545: Futur
eWarning: Passing (type, 1) or '1type' as a synonym of type is deprecated; in a future version of numpy, it will be understood
as (type, (1,)) / '(1,)type'.
  _np_qint32 = np.dtype([("qint32", np.int32, 1)])
c:\users\admin\appdata\local\programs\python\python37\lib\site-packages\tensorboard\compat\tensorflow_stub\dtypes.py:550: Futur
eWarning: Passing (type, 1) or '1type' as a synonym of type is deprecated; in a future version of numpy, it will be understood
as (type, (1,)) / '(1,)type'.
  np_resource = np.dtype([("resource", np.ubyte, 1)])
```

```
In [2]: #define and load class labels found in dataset
path = "Dataset"
labels = []
X = []
Y = []
for root, dirs, directory in os.walk(path):
    for j in range(len(directory)):
        name = os.path.basename(root)
        if name not in labels:
            labels.append(name.strip())
print("Chest Disease Class Labels : "+str(labels))
```

```
Chest Disease Class Labels : ['Fibrosis', 'Inflammation', 'Normal', 'Pneumonia']
```

```
In [3]: #define function to get class label of given image
```

Fig.2. Upload Operation



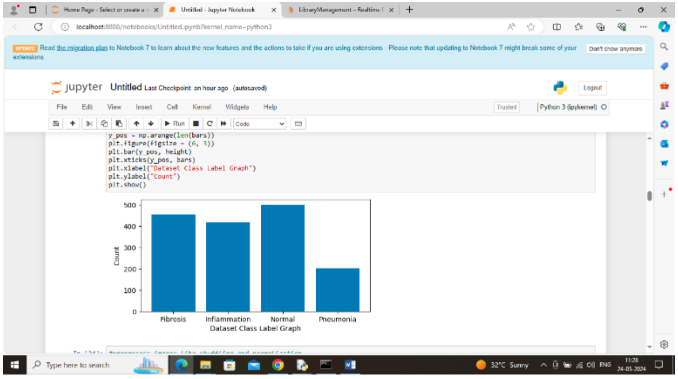
```
    X = np.load('model/X.txt.npy')
    Y = np.load('model/Y.txt.npy')
else: #if images not process then read and process image pixels
    for root, dirs, directory in os.walk(path):#connect to dataset folder
        for j in range(len(directory)):#loop all images from dataset folder
            name = os.path.basename(root)
            if 'Thumbs.db' not in directory[j]:
                img = cv2.imread(root+"/"+directory[j])#read images
                img = cv2.resize(img, (32, 32))#resize image
                X.append(img) #add image pixels to X array
                label = getLabel(name)#get image label id
                Y.append(label)#add image label
    X = np.asarray(X)#convert array as numpy array
    Y = np.asarray(Y)
    np.save('model/X.txt',X)#save process images and labels
    np.save('model/Y.txt',Y)
print("Dataset images loaded")
print("Total images found in dataset : "+str(X.shape[0]))
print()
```

```
Dataset images loaded
Total images found in dataset : 1572
```

```
In [23]: #visualizing class labels count found in dataset
names, count = np.unique(Y, return_counts = True)
```

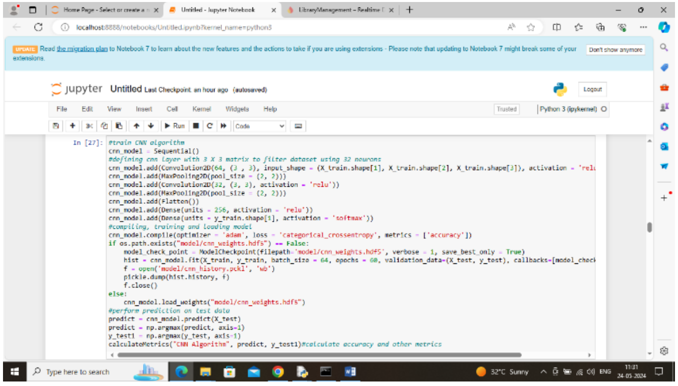
Fig.3. Healthcare Dataset

In above screen connecting to dataset folder and then looping and reading all images and then resizing to equal size and then creating X and Y training array where X will contains image features and Y will contains labels and then in blue color text can see total number of images



```
y_pos = np.arange(len(bars))
plt.figure(figsize = (6, 3))
plt.bar(y_pos, height)
plt.xticks(y_pos, bars)
plt.xlabel("Dataset Class Label Graph")
plt.ylabel("Count")
plt.show()
```

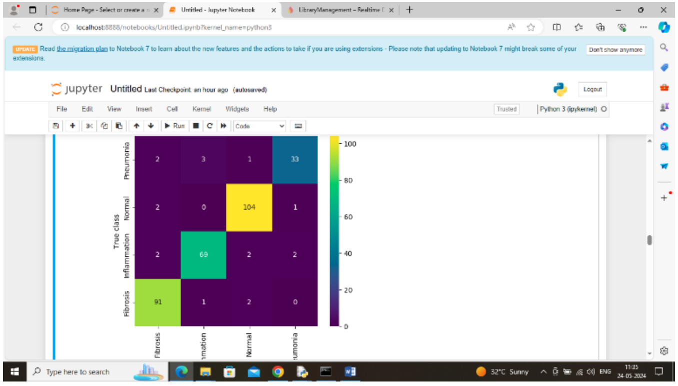
Fig.4. Dataset Class Label Graph



```
In [27]: #train CNN algorithm
cnn_model = Sequential()
#defining cnn layer with 3 X 3 matrix to filter dataset using 32 neurons
cnn_model.add(Convolution2D(64, (3 , 3), input_shape = (X_train.shape[1], X_train.shape[2], X_train.shape[3]), activation = 'relu
cnn_model.add(MaxPooling2D(pool_size = (2, 2)))
cnn_model.add(Convolution2D(32, (3, 3), activation = 'relu'))
cnn_model.add(MaxPooling2D(pool_size = (2, 2)))
cnn_model.add(Flatten())
cnn_model.add(Dense(units = 256, activation = 'relu'))
cnn_model.add(Dense(units = y_train.shape[1], activation = 'softmax'))
#compiling, training and loading model
cnn_model.compile(optimizer = 'adam', loss = 'categorical_crossentropy', metrics = ['accuracy'])
if os.path.exists("model/cnn_weights.hdf5") == False:
    model_check_point = ModelCheckpoint(filepath='model/cnn_weights.hdf5', verbose = 1, save_best_only = True)
    hist = cnn_model.fit(X_train, y_train, batch_size = 64, epochs = 60, validation_data=(X_test, y_test), callbacks=[model_check
    f = open('model/cnn_history.pckl', 'wb')
    pickle.dump(hist.history, f)
    f.close()
else:
    cnn_model.load_weights("model/cnn_weights.hdf5")
#perform prediction on test data
predict = cnn_model.predict(X_test)
predict = np.argmax(predict, axis=1)
y_test1 = np.argmax(y_test, axis=1)
calculateMetrics("CNN Algorithm", predict, y_test1)#calculate accuracy and other metrics
```

In above screen defining CNN algorithm layers for features filtration and training and then generate a train model and this model will be applied on test data to get below output

Fig.5. Train CNN Algorithm



In above confusion matrix graph x-axis represents 'Predicted Labels' and y-axis represents true labels and then all different color boxes in diagnol

Fig.6. CNN Algorithm Confusion Matrix

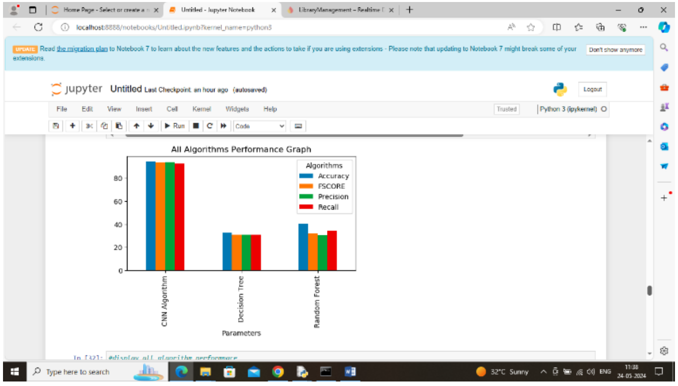
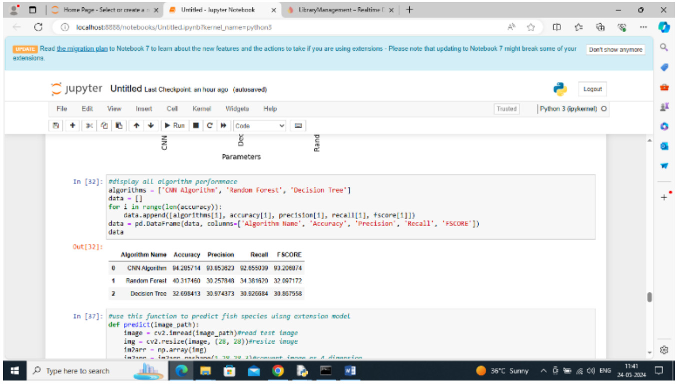


Fig.7. Parameter Matrix

in above graph showing comparison between all algorithms where x-axis represents algorithm names and y-axis represents accuracy and other metrics in different color bars and in all algorithms CNN got best performance



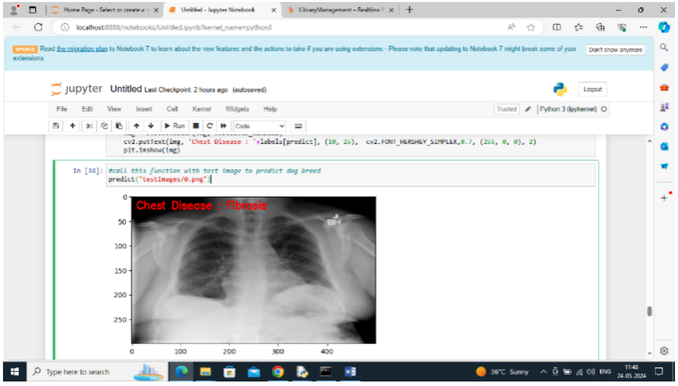
```
In [32]: #display all algorithm performnace
         algorithms = ['CNN Algorithm', 'Random Forest', 'Decision Tree']
         data = []
         for i in range(len(accuracy)):
             data.append([algorithms[i], accuracy[i], precision[i], recall[i], fscore[i]])
         data = pd.DataFrame(data, columns=['Algorithm Name', 'Accuracy', 'Precision', 'Recall', 'FSCORE'])
         data

Out[32]:
      Algorithm Name   Accuracy  Precision     Recall     FSCORE
   0   CNN Algorithm  94.285714  93.853623  92.855039  93.206874
   1   Random Forest  40.317460  30.257848  34.381620  32.097172
   2   Decision Tree  32.698413  30.974373  30.926684  30.867558

In [37]: #use this function to predict fish species uisng extension model
         def predict(image_path):
             image = cv2.imread(image_path)#read test image
             img = cv2.resize(image, (28, 28))#resize image
             im2arr = np.array(img)
```

Fig.8 Comparision Graph

In above screen displaying all algorithm performance in tabular format



```
    cv2.putText(img, 'Chest Disease : '+labels[predict], (10, 25),  cv2.FONT_HERSHEY_SIMPLEX,0.7, (255, 0, 0), 2)
    plt.imshow(img)

In [38]: #call this function with test image to predict dog breed
         predict("testImages/0.png")
```

Fig.9. Pneumonia Result

In above screen calling predict function with image path and then CNN detected 'Fibrosis' from given image

5. CONCLUSION

In this paper, we proposed a novel approach composed of two steps for chest disease classifcation using new DL architectures. In the frst step (multi-classifcation), we classifed chest X-ray (CXR) images into three classes (normal, lung disease, and heart disease). In the second step (binary classifcation), we classifed CXR images into specifc diseases to predict whether it is a normal or abnormal case. A dataset of 26,316 CXR images was consolidated by merging images from two public datasets (VinDr-CXR and CheXpert) to train, validate, and test our methods. For this work, we implemented two DL methods to perform our two-step classifcation approach.

REFERENCES

[1]. Ambati A, Dubey SR. Ac-covidnet: Attention guided contrastive cnn for recognition of covid-19 in chest x-ray images. In: International Conference on Computer Vision and Image Processing, pp. 71–82 (2022).

[2]. Russo P. Handbook of X-ray Imaging: Physics and Technology. Series in Medical Physics and Biomedical Engineering. CRC Press, New York (2017).

[3]. World Health Organization: To X-ray or not to X-ray. https://www. who.int/news-room/feature-stories/detail/to-x-ray-or-not-to-x-ray- (Accessed: November 2022).

[4]. Nasser AA, Akhlouf M. Chest diseases classifcation using cxr and deep ensemble learning. In: 19th International Conference on Content-based Multimedia Indexing, pp. 116–120 (2022).

[5]. Marciniuk DD, Schraufnage DE, Ferkol T, Fong KM, Joos G, Varela VL, Zar H. The Global Impact of Respiratory Disease. European Respiratory Society, ??? (2017) .

[6]. Tsuneki M. Deep learning models in medical image analysis. J Oral Biosci. 2022;64(3):312–20.

[7]. Demner-Fushman D, Kohli MD, Rosenman MB, Shooshan SE, Rodriguez L, Antani S, Thoma GR, McDonald CJ. Preparing a collection of radiology examinations for distribution and retrieval. J Am Med Inform Assoc. 2016;23(2):304–10.

[8]. Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM. Chestxray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classifcation and localization of common thorax diseases. In: IEEE Conference on Computer Vision and Pattern Reconition (CVPR), 2017; pp. 2097–2106

[9]. Nguyen H, Lam K, Le L, Pham H, Tran D, Nguyen D, Le D, Pham C, Tong H, Dinh D, et al. Vindr-cxr: An open dataset of chest x-rays with radiologist's annotations. 2021; preprint arXiv:2012. 15029

[10]. Irvin J, Rajpurkar P, Ko M, Yu Y, Ciurea-Ilcus S, Chute C, Marklund H, Haghgoo B, Ball R, Shpanskaya K, Seekins J, Mong D, Halabi S, Sandberg J, Jones R, Larson D, Langlotz C, Patel B, Lungren M, Ng A. Chexpert : A large chest radiograph dataset with uncertainty labels and expert comparison. In: AAAI Conference on Artifcial Intelligence, 2019; pp. 590–597.

[11]. Zhu CS, Pinsky PF, Kramer BS, Prorok PC, Purdue MP, Berg CD, Gohagan JK. The prostate, lung, colorectal, and ovarian cancer screening trial and its associated research resource. JNCI. 2013;105(22):1684–93.

[12]. Johnson AE, Pollard TJ, Berkowitz SJ, Greenbaum NR, Lungren MP, Deng C-Y, Mark RG, Horng S. Mimic-cxr, a de-identifed publicly available database of chest radiographs with free-text reports. Sci Data. 2019;6(1):1–8.

[13]. Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, Ding DY, Bagul A, Langlotz CP, Shpanskaya KS, Lungren MP, Ng AY. Chexnet: Radiologist-level pneumonia detection on chest X-rays with deep learning. 2017; arXiv preprint arXiv:1711.05225

[14]. Sze-To A, Wang Z. tchexnet: Detecting pneumothorax on chest x-ray images using deep transfer learning. In: International Conference on Image Analysis and Recognition, 2019; pp. 325–332.

[15]. Khoiriyah SA, Basof A, Fariza A. Convolutional neural network for automatic pneumonia detection in chest radiography. In: International Electronics Symposium (IES), 2020; pp. 476–480

[16]. Wang H, Xia Y. Chestnet: A deep neural network for classifcation of thoracic diseases on chest radiography. 2018; arXiv preprint arXiv:1807.03058

[17]. Pham HH, Le TT, Tran DQ, Ngo DT, Nguyen HQ. Interpreting chest X-rays via cnns that exploit hierarchical disease dependencies and uncertainty labels. Neurocomputing. 2021;437:186–94.

[18]. Guendel S, Grbic S, Georgescu B, Liu S, Maier A, Comaniciu D. Learning to recognize abnormalities in chest X-rays with locationaware dense networks. In: Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, 2019; pp. 757–765.

[19]. Kim S, Rim B, Choi S, Lee A, Min S, Hong M. Deep learning in multi-class lung diseases' classifcation on chest X-ray images. Diagnostics. 2022;12(4):915.

[20]. Blais M-A, Akhlouf MA. Deep learning and binary relevance classifcation of multiple diseases using chest X-ray images. In: 43rd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC), 2021; pp. 2794–2797.

[1] Arce, M.E.; Crespo, B.; Miguel-Álvarez, C. Higher Education Drop-out in Spain–Particular Case of Universities in Galicia. Int. Educ. Stud. 2015, 8, 247–264.

[2] Xavier, M.; Meneses, J. Dropout in Online Higher Education: A Scoping Review from 2014 to 2018; eLearn Centre, Universität Oberta de Catalunya: Barcelona, Spain, 2020.

[3] Baker, R.S.; Invent ado, P.S. Educational Data Mining and Learning Analytics. In Learning Analytics: From Research to Practice; Larusson, J.A., White, B., Eds.; Springer: New York, NY, USA, 2014; pp. 61–75.

[4] Moreno-Marcos, P.M.; Alario-Hoyos, C.; Maoz-Merino, P.J.; Kloos, C.D. Prediction in MOOCs: A Review and Future Research Directions. IEEE Trans. Learn. Technol. 2019, 12, 384–401.

[5] Ranjeeth, S.; Patchouli, T.; Paul, P.V. A survey on predictive models of learning analytics. Procedia Computer. Sci. 2020, 167, 37–46.

[6] Hamim, T.; Benab Bou, F.; Sael, N. Survey of Machine Learning Techniques for Student Profile Modelling. Int. J. Emerg. Technol. Learn. 2021, 16, 136–151.

[7] Prianka, B.; Velardi, P.; Stilo, G.; Distant, D.; Ferally, S. A survey of machine learning approaches for student dropout prediction in online courses. ACM Computer. Surd. (CSUR) 2020, 53, 1–34.

[8] Gardner, J.; Brooks, C. Student success prediction in MOOCs. User Model. User-Adapt. Interact. 2018, 28, 127–203.

[9] Katara, R.; Gaba, J.; Garg, A.; Verma, V. A review on machine learning based student's academic performance prediction systems. In Proceedings of the 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), Coimbatore, India, 25–27 March 2021; pp. 254–259.

[10] Filius, R.M.; Bijl, S.G. Teaching Methodologies for Scalable Online Education. In Handbook for Online Learning Contexts: Digital, Mobile and Open; Springer: Berlin/Heidelberg, Germany, 2021; pp. 55–65.

[11] Amirah, E.A.; Hamsini, T.M.; Ajara, I. Mining Educational Data to Predict Student's academic Performance using Ensemble Methods. Int. J. Database Theory Appl. 2016, 9, 119–136.

[12] Rahman, M.H.; Islam, M.R. Predict Student's Academic Performance and Evaluate the Impact of Different Attributes on the Performance Using Data Mining Techniques. In Proceedings of the 2017 2nd International Conference on Electrical Electronic Engineering (ICEEE), Rajdhani, Bangladesh, 27–29 December 2017; pp. 1–4.

[13] Kozelek, J.; Hosta, M.; Rahal, Z. Open university learning analytics dataset. Sci. Data 2017, 4, 170171. [

[14] Adnan, M.; Habib, A.; Ashraf, J.; Mus Sadiq, S.; Raza, A.A.; Abid, M.; Bashir, M.; Khan, S.U. Predicting at-Risk Students at Different Percentages of Course Length for Early Intervention Using Machine Learning Models. IEEE Access 2021, 9, 7519–7539.

[15] Jha, N.; Georgescu, I.; Moldovan, A. OULAD MOOC Dropout and Result Prediction using Ensemble, Deep Learning and Regression Techniques. In Proceedings of the 11th International Conference on Computer Supported Education, Heraklion, Greece, 2–4 May 2019; Scoters: Setubal, Portugal, 2019; Volume 2, pp. 154–164.

[16] Song, X.; Li, J.; Sun, S.; Yin, H.; Dawson, P.; Doss, R.R.M. SEPN: A Sequential Engagement Based Academic Performance Prediction Model. IEEE Intel. Syst. 2021, 36, 46–53.

[17] Hussain, M.; Zhu, W.; Zhang, W.; Abidi, R. Student Engagement Predictions in an e-Learning System and Their Impact on Student Course Assessment Scores. Computer. Intel. Neurosis. 2018, 2018, 6347186.

[18] Stanford, U. Centre for Advanced Research through Online Learning (CAROL). Available online,

[19] Mubarak, A.A.; Cao, H.; Hizam, I.M. Deep analytic model for student dropout prediction in massive open online courses. Computer. Electra. Eng. 2021, 93, 107271.

[20] Mouridi, Y.; Saigal, M.; El Katine, H.; Berrada Fathi, W. A machine learning-based methodology to predict learners' dropout, success or failure in MOOCs. Int. J. Web Inf. Syst. 2019, 15, 489–509.

[21] Mouridi, Y.; Saigal, M.; Berrada Fathi, W.; El Katine, H. A Machine Learning Based Approach to Enhance Mooch Users' Classification. Turk. Online J. Distance Educ. 2020, 21, 47–68.

[22] KDD. KDD Cup 2015. Available online.

[23] Lai, S.; Zhao, Y.; Yang, Y. Broad Learning System for Predicting Student Dropout in Massive Open Online Courses. In Proceedings of the 2020 8th International Conference on Information and Education Technology, Okayama, Japan, 28–30 March 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 12–17.

[24] Zhou, Y.; Xu, Z. Multi-Model Stacking Ensemble Learning for Dropout Prediction in MOOCs. J. Phys. Conf. Ser. 2020, 1607, 012004.

[25] Ho, A.D.; Reich, J.; Nesterov, S.; Seaton, D.T.; Mullaney, T.; Waldo, J.; Chuang, I. Harvard and MITx: The first year of Open Online Courses; Harvard and MITx Working Paper No. 1; Harvard University: Cambridge, MA, USA, 2014.

[26] Imran, A.; Dalipi, F.; Kastrati, Z. Predicting Student Dropout in a MOOC: An Evaluation of a Deep Neural Network Model. In Proceedings of the 2019 5th International Conference on Computing and Artificial Intelligence, Bali, Indonesia, 19–22 April 2019; pp. 190–195.

[27] Al-Sha bandar, R.; Hussain, A.; Laws, A.; Keight, R.; Lunn, J.; Radi, N. Machine learning approaches to predict learning outcomes in Massive open online courses. In Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA, 14–19 May 2017; pp. 713–720.

[28] Liu, K.F.-R.; Chen, J.-S. Prediction and assessment of student learning outcomes in calculus a decision support of integrating data mining and Bayesian belief networks. In Proceedings of the 2011 3rd International Conference on Computer Research and Development, Shanghai, China, 11–13 March 2011; Volume 1, pp. 299–303.

[29] Wu, W.H.; Jim Wu, Y.C.; Chen, C.Y.; Kao, H.Y.; Lin, C.H.; Huang, S.H. Review of Trends from Mobile Learning Studies: A Meta-Analysis. Computer. Educ. 2012, 59, 817–827.

[30] Chen, W.; Brinton, C.G.; Cao, D.; Mason-Singh, A.; Lu, C.; Chiang, M. Early Detection Prediction of Learning Outcomes in Online Short-Courses via Learning Behaviours. IEEE Trans. Learn. Technol. 2019, 12, 44–58.