﷽

**SLIDE**

**Developing a Pneumonias Disease Prediction System using CNN-based Deep Neural Network**

**(Machine Learning – Semester Project)**

**Presented by:**

**Ms. Fatima Zulfiqar SP20-RCS-005**

**Mr. Rehan Raza SP20-RCS-018**

**Course Taught by:**

**Dr. Rao Muhammad Adeel Nawab**

**Department of Computer Science,**

**Comsats University Islamabad, Lahore Campus**

**SLIDE**

**Outline**

* **Introduction**
* **Proposed Technique**
* **Experimental Setup**
* **Results and Analysis**
* **Conclusion and Future Work**

**SLIDE**

**===========**

**Introduction**

**===========**

**SLIDE**

**Pneumonia Diseases – Definition**

* **Pneumonia is a very infective disease that is caused by bacteria, viruses, or fungi**

**SLIDE**

**Pneumonia Disease – World Wide Death Rate**

* **According to a recent survey conducted by World Health Organization(WHO),** 
  + **Every year 700,000 children die because of Pneumonia Disease in the World**

**SLIDE**

**Pneumonia Disease – Advantages of Early Stage Diagnose**

* **Early Stage Diagnosis of Pneumonia Disease can help**
  + **Save many lives**
* **Accurate Predictions Systems can help us** 
  + **Diagnose Pneumonia Disease (from X-ray Images) quickly and efficiently**

**SLIDE**

**Semester Project Focus**

* **Develop a Pneumonia Disease Prediction (from X-ray Images) System using CNN-based Deep Neural Networks**

**SLIDE**

**Pneumonia Disease Prediction (from X-ray Image) Problem – Given and Task**

* **Given**

1. **An X-ray Image**
2. **Finite Set of Labels (Normal and Pneumonia)**

* **Task**
  + **Assign a Single Label to the given X-ray Image from the pre-defined Set of Labels**

**SLIDE**

**Pneumonia Disease Prediction (from X-ray Image) Problem – Input and Output**

* **Input**
  + **An X-ray Image (Unstructured)**
* **Output**
  + **Normal / Pneumonia (Categorical)**

**SLIDE**

**Example 01 – Pneumonia Disease Prediction (from X-ray Image) System**

* **Input**

|  |
| --- |
|  |

* **Output**

|  |
| --- |
| **Normal** |

**SLIDE**

**Example 02 – Pneumonia Disease Prediction (from X-ray Image) System**

* **Input**

|  |
| --- |
|  |

* **Output**

|  |
| --- |
| **Pneumonia** |

**SLIDE**

**=================**

**Proposed Technique**

**=================**

**SLIDE**

**Baseline Technique**

* **As Baseline Technique, we used**
  + **Majority Class Categorization**
* **Majority Class Categorization is calculated by assigning the label of Majority Class to all the Test Instances**
* **Formula**

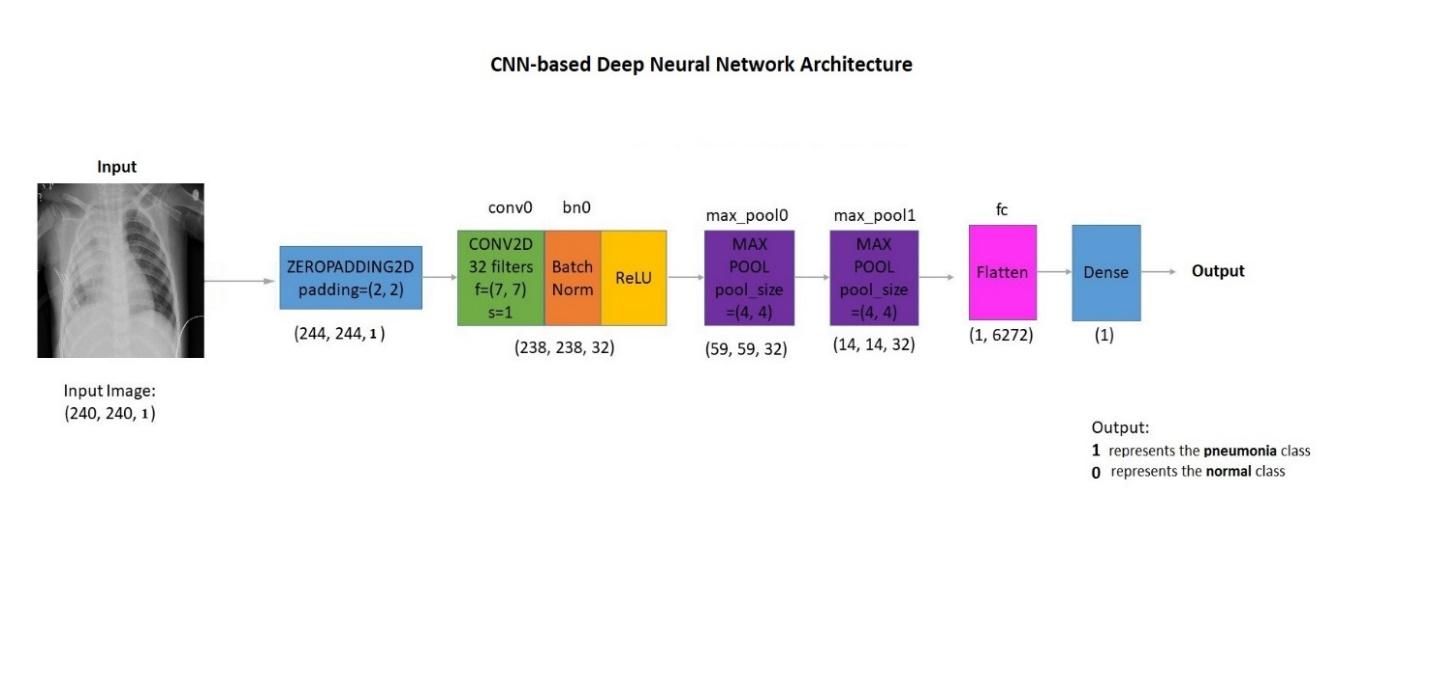
**SLIDE**

**Proposed Technique**

* **To develop our proposed Pneumonia Disease Prediction (from X-ray Image) System, we used**
  + **CNN-based Deep Neural Network**

**SLIDE**

**Architecture - CNN-based Deep Neural Network**



**SLIDE**

**Main Parameters - CNN-based Deep Neural Networks**

* **For CNN-based Deep Neural Network used in this Problem, following main Parameters were used**
  + **No. of Input Layer = 1**
    - **Input Size = (224 X 224 X 1)**
  + **No. of Hidden Layer = 2**
    - **No. of Convolutional Layer = 1**
    - **No. of Pooling Layer = 2**
    - **No. of Batch Normalization Layer = 1**
    - **Flatten Layer = 1**
  + **No. of Output Layer = 1**
    - **No. of Output Unit = 1**
  + **Activation Function = 1**
    - **Hidden Layer = ReLU**
    - **Output Layer = Sigmoid**
  + **Number of Epochs = 15**
  + **Learning Rate = 0.0001**
  + **Optimizer = Adam**
  + **Loss Function = binary\_crossentropy**

**SLIDE**

**=================**

**Experimental Setup**

**=================**

**SLIDE**

**Experimental Setup**

* **The Experimental Setup comprises the following four main components**
  + **Dataset**
  + **Techniques**
  + **Evaluation Methodology**
  + **Evaluation Measures**

**-----------**

**Dataset**

**-----------**

**SLIDE**

**Dataset**

* **Dataset Name**
  + **Pneumonia Disease Dataset**
* **Dataset Download Link** 
  + [**https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia**](https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia)

**SLIDE**

**Dataset - Main Characteristics**

|  |  |
| --- | --- |
| Class | Number of Instances |
| Normal | **1,583 (27%)** |
| Pneumonia | **4,273 (73%)** |
| Total | **5,856** |

**SLIDE**

**Data Split**

* **We Split Pneumonia Disease Dataset using Class Balanced Data Split Approach** 
  + **Training Data = 80% (4,684 Instances)**
  + **Testing Data = 10% (587 Instances)**
  + **Validation Data = 10% (585 Instances)**

|  |  |  |  |
| --- | --- | --- | --- |
| Class | Training Set | Testing Set | Validation Set |
| Normal | 1,266 | 159 | 158 |
| Pneumonia | 3,418 | 428 | 427 |
| Total | **4,684** | **587** | **585** |

**---------------**

**Techniques**

**---------------**

**SLIDE**

**Techniques**

* **Baseline Technique** 
  + **Majority Class Categorization**
* **Proposed Technique** 
  + **CNN-based Deep Neural Network**

**SLIDE**

**Parameters Settings – CNN-based Deep Neural Network**

|  |  |
| --- | --- |
| Parameters Settings – CNN-based Deep Neural Network | |
| Parameter | **Parameter Value** |
| Number of Epoch | 15 |
| Optimizer | Adam |
| Learning Rate | 0.0001 |
| Loss Function | binary\_crossentropy |

**SLIDE**

**Implementation - CNN-based Deep Neural Network**

|  |  |
| --- | --- |
| System Settings | |
| Developer Name | Ms. Fatima Zulfiqar & Mr. Rehan Raza |
| Programming Language | Python 3.7.4 |
| IDE | Google Colab Notebook |
| Deep Learning Toolkit | Tensorflow 2.0  Keras 2.2 |
| Code Version | 1.0 |
| Date | 01 – December – 2020 |

**----------------------------------**

**Evaluation Methodology**

**----------------------------------**

**SLIDE**

**Evaluation Methodology**

* **Pneumonia Disease Problem** 
  + **Treated as Binary Classification Problem**
* **Main Aim** 
  + **Differentiate between two Classes**
    - **Normal**
    - **Pneumonia**
* **Data Split Approach** 
  + **Training Data = 80% (4,684 Instances)**
  + **Testing Data = 10% (587 Instances)**
  + **Validation Data = 10% of Training Data (585 Instances)**
* **Training Data** 
  + **Used to Build the Model**
* **Validation Data**
  + **Used to check whether the Model is Overfitting during Training or Not?**
* **Testing Data**
  + **Used to evaluate the performance of the (Trained) Model**

**-----------------------------**

**Evaluation Measures**

**-----------------------------**

**SLIDE**

**Evaluation Measure**

* **Evaluation was carried out using** 
  + **Accuracy**
  + **Precision**
  + **Recall**
  + **F1**
* **Accuracy is defined as the proportion of correctly classified Test instances**
* **Precision (P) is defined as the proportion of the predicted Positive cases that were correct**
* **Recall is defined as the proportion of Positive cases that were correctly classified**
* **F1 – measure is defined as Harmonic mean of Precision and Recall, When we assign same weights to Precision and Recall the F – measure become F1 – measure**

**=================**

**Results and Analysis**

**=================**

**SLIDE**

**Results using Most Common Categorization**

* **Total Instances in Test Data = 587**
  + **Normal = 159**
  + **Pneumonia = 428**
* **Applying MCC on Test Data**

**SLIDE**

**Results and Analysis**

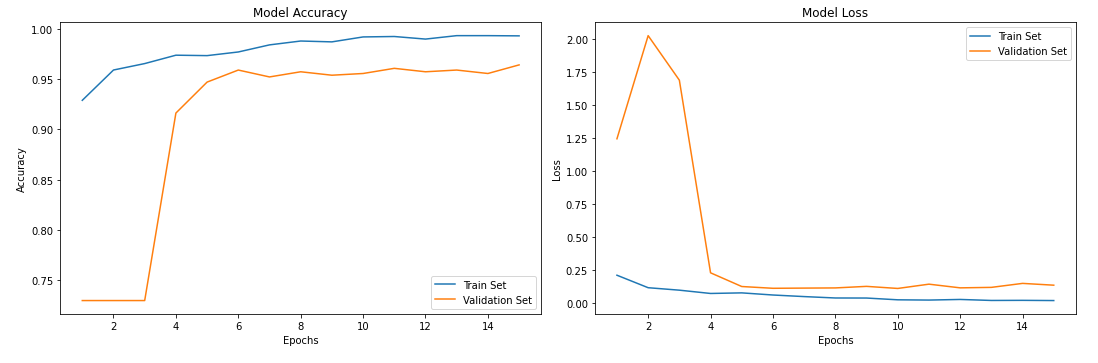
* **Weighted Average Precision, Recall and F1 scores are reported in Table below**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Technique | Class | Accuracy | Precision | Recall | F1 – Score |
| MCC | **Normal** | - | - | - | - |
|  | **Pneumonia** | - | - | - | - |
|  | **Macro Average** | 0.73 | - | - | - |
| CNN | **Normal** | 0.99 | 0.91 | 0.99 | 0.95 |
|  | **Pneumonia** | 0.96 | 1.00 | 0.96 | 0.98 |
|  | **Macro Average** | 0.98 | 0.95 | 0.98 | 0.96 |

**SLIDE**

**Detailed Analysis - Proposed CNN-based Deep Neural Network**

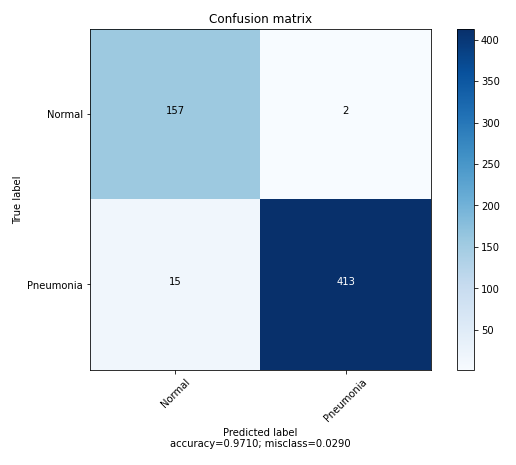
* **Training-Validation Curve**



**SLIDE**

**Detailed Analysis - Proposed CNN-based Deep Neural Network**

* **Confusion Matrix**



**SLIDE**

**========================**

**Conclusion and Future Work**

**========================**

**SLIDE**

**Main Findings**

* **Finding 01**
  + **Proposed Technique (CNN Model) outperforms the Baseline Technique (MCC)**
* **Finding 02**
  + **Accuracy of CNN Model is 0.97 compared to 0.73 obtained with MCC**
    - **This indicates a huge improvement in performance**
* **Finding 03**
  + **The most problematic Class is Pneumonia**
* **Finding 04**
  + **Google Colab is a good platform to run experiments (on large Data) particularly using CNN-based Deep Neural Networks**

**SLIDE**

**Conclusion**

* **Proposed CNN-based Model performs very well in predicting Pneumonia Disease (from X-ray Images) with an Accuracy of 97%**

**SLIDE**

**Future Work**

* **In Future, we plan to**
  + **Increase Dataset Size**
  + **Explore Data Augmentation**
  + **Explore other Techniques including** 
    - **Capsule Neural Network**
    - **Transfer Learning**