PNEUMONIA DETECTION USING ENSEMBLE LEARNING

Anitha Moses V[*1],Akshaya K[*2], Anusha R[*3], Harrsheetha S[*4]

[*1]Professor, Department of CSE, Panimalar Engineering College, Chennai, TamilNadu, India
[*2]Student, Department of CSE, Panimalar Engineering College, Chennai, TamilNadu, India
[*3]Student, Department of CSE, Panimalar Engineering College, Chennai, TamilNadu, India
[*4]Student, Department of CSE, Panimalar Engineering College, Chennai, TamilNadu, India

ABSTRACT:

Pneumonia is a life-threatening infectious disease affecting one or both lungs in humans commonly caused by bacteria called Streptococcus pneumonia. 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, Convolution 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.

INTRODUCTION:

Generally, data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases. While large-scale information technology has been evolving separate transaction and analytical systems, data mining provides the link between the two. Data mining software analyzes relationships and patterns in stored transaction data based on open-ended user queries. Several types of analytical software are available: statistical, machine learning, and neural networks.

Generally, any of four types of relationships are sought: 1) Classes: Stored data is used to locate data in predetermined groups. For example, a restaurant chain could mine customer purchase data to determine when customers visit and what they typically order. This information could be used to increase traffic by having daily specials. 2) Clusters: Data items are grouped according to logical relationships or consumer preferences. For example, data can be mined to identify market segments or consumer affinities. 3) Associations: Data can be mined to identify associations. The beer-diaper example is an example of associative mining. 4) Sequential patterns: Data is mined to anticipate behavior patterns and trends. For example, an outdoor equipment retailer could predict the likelihood of a backpack being purchased based on a consumer's purchase of sleeping bags and hiking shoes.

Data mining consists of five major elements: 1) Extract, transform, and load transaction data onto the data warehouse system. 2) Store and manage the data in a multidimensional database system. 3) Provide data access to business analysts and information technology professionals. 4) Analyze the data by application software. 5) Present the data in a useful format, such as a graph or table. Different levels of analysis are available: 1) Artificial neural networks: Non-linear predictive models that learn through training and resemble biological neural networks in structure. 2) Genetic algorithms: Optimization techniques that use process such as genetic combination, mutation, and natural selection in a design based on the concepts of natural evolution. 3) Decision trees: Tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset.

Specific decision tree methods include Classification and Regression Trees (CART) and Chi Square Automatic Interaction Detection (CHAID). CART and CHAID are decision tree techniques used for classification of a dataset. They provide a set of rules that you can apply to a new (unclassified) dataset to predict which records will have a given outcome. CART segments a dataset by creating 2-way splits while CHAID segments using chi square tests to create multi-way splits. CART typically requires less data preparation than CHAID. 4) Nearest neighbor method: A technique that classifies each record in a dataset based on a combination of the classes of the k record(s) most similar to it in a historical dataset (where k=1). Sometimes called the k-nearest neighbor technique. 5) Rule induction: The extraction of useful if-then rules from data based on statistical significance. 6) Data visualization: The visual interpretation of complex relationships in multidimensional data. Graphics tools are used to illustrate data relationships.

Characteristics of Data Mining: 1) Large quantities of data: The volume of data so great it has to be analyzed by automated techniques e.g. satellite information, credit card transactions etc. 2) Noisy, incomplete data: Imprecise data is the characteristic of all data collection. 3) Complex data structure: conventional statistical analysis not possible 4) Heterogeneous data stored in legacy systems. Benefits of Data Mining: 1) it's one of the most effective services that are available today. With the help of data mining, one can discover precious information about the customers and their behavior for a specific set of products and evaluate and analyze, store, mine and load data related to them 2)An analytical CRM model and strategic business related decisions can be made with the help of data mining as it helps in providing a complete synopsis of customers 3)An endless number of organizations have installed data mining projects and it has helped them see their own companies make an unprecedented improvement in their marketing strategies (Campaigns) 4) Data mining is generally used by organizations with a solid customer focus. For its flexible nature as far as applicability is concerned is being used vehemently in applications to foresee crucial data including industry analysis and consumer buying behaviors 5) Fast paced and prompt access to data along with economic processing techniques have made data mining one of the most suitable services that a company seek. Advantages of Data Mining: 1)Marketing / Retail: Data mining helps marketing companies build models based on historical data to predict who will respond to the new marketing campaigns such as direct mail, online marketing campaign…etc. Through the results, marketers will have appropriate approach to sell profitable products to targeted customers.

Data mining brings a lot of benefits to retail companies in the same way as marketing. Through market basket analysis, a store can have an appropriate production arrangement in a way that customers can buy frequent buying products together with pleasant. In addition, it also helps the retail companies offer certain discounts for particular products that will attract more customers. 2) Finance / Banking: Data mining gives

financial institutions information about loan information and credit reporting. By building a model from historical customer's data, the bank and financial institution can determine good and bad loans. In addition, data mining helps banks detect fraudulent credit card transactions to protect credit card's owner. 3) Manufacturing : By applying data mining in operational engineering data, manufacturers can detect faulty equipments and determine optimal control parameters. For example semi-conductor manufacturers has a challenge that even the conditions of manufacturing environments at different wafer production plants are similar, the quality of wafer are lot the same and some for unknown reasons even has defects.

Data mining has been applying to determine the ranges of control parameters that lead to the production of golden wafer. Then those optimal control parameters are used to manufacture wafers with desired quality. 4) Governments: Data mining helps government agency by digging and analyzing records of financial transaction to build patterns that can detect money laundering or criminal activities. 5) Law enforcement:: Data mining can aid law enforcers in identifying criminal suspects as well as apprehending these criminals by examining trends in location, crime type, habit, and other patterns of behaviors. 6) Researchers: Data mining can assist researchers by speeding up their data analyzing process; thus, allowing those more time to work on other projects.

EXISTING SYSTEM:

Antibiotics are used to recover the bacterial infected pneumonia while viral infected patient need the different medication and supportive care for recovery of the diseases. Chest X-ray (CXR) analysis is the chosen as one of the most preferably for the medical practitioner for diagnosing the pneumonia disease. Accurately identifying and categorizing the pneumonia subtypes is an important and challenging clinical task, and automated methods can be used to save time and reduce error. Deep learning techniques can be used to recognize and identify the type of pneumonia infection in the patient and proper clinical treatment can be provided towards making the patient cure from this kind of respiratory infection. Deep Siamese based CNN method can be associated with the radiographic examination towards detecting the particular type of pneumonia.

DISADVANTAGES OF EXISTING SYSTEM:

Despite of several advantages of X-ray imaging, still in some cases it is not possible to identify the correct region of interest in the radiographic image for detecting the diseases. It is also observed that as the diagnostic accuracy of automated method reach the human level. This makes researchers to develop automated methods for detection of diseases through radiography imagery. Automated detection method provides an add on support to the clinical experts towards an ease of diagnosing the disease.

PROPOSED SYSTEM:

An optimum solution for the detection of pneumonia from chest X-rays is proposed in this work. Data augmentation was used to address the problem of the limited dataset, and then, state-of-the-art deep learning models, as discussed and was fine-tuned for pneumonia classification. Then, predictions from these models were combined, using a weighted classifier to compute the final prediction.

ADVANTAGES OF PROPOSED SYSTEM:

To sum up, our work has three major contributions as follows: have trained our model for both the right and left part of chest X-ray that makes the process robust towards better classification. Overall our approach has provided the add on benefits to inexperienced practitioners in towards diagnosing the specific type of pneumonia in the early stage.

SYSTEM ARCHITECTURE:
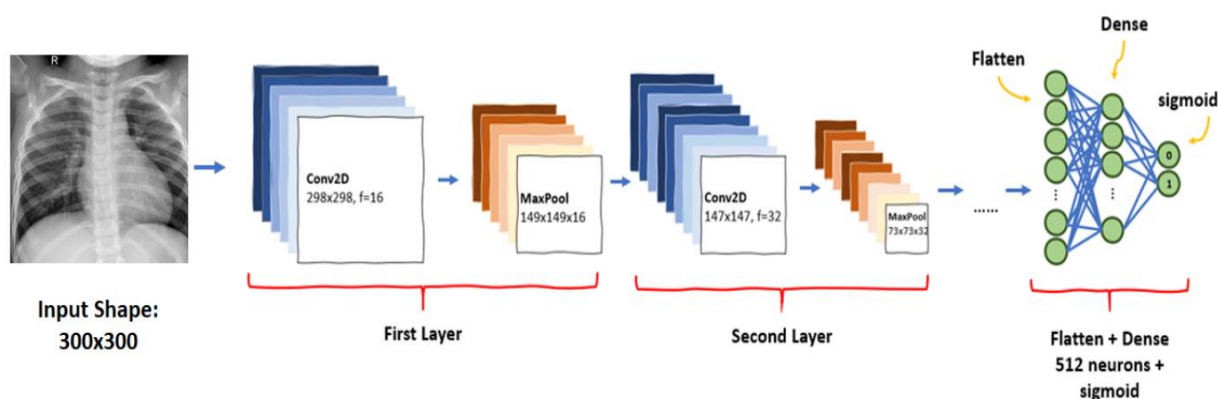
The modules involved in this project:

- Dataset Collection
- Data Preprocessing
- Splitting training and testing data
- Build CNN VGG16 AND RESNET50 model
- Testing and evaluate the model

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy. Input Design considered the following things: What data should be given as input? How the data should be arranged or coded? The dialog to guide the operating personnel in providing input. Methods for preparing input validations and steps to follow when error occur. A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system's relationship to help user decision-making. 1. Designing computer output should proceed in an organized, well thought out manner; the right output must be developed while ensuring that each output element is designed so that people will find the system can use easily and effectively. When analysis design computer output, they should Identify the specific output that is needed to meet the requirements. 2. Select methods for presenting information. 3. Create document, report, or other formats that contain information produced by the system. The output form of an information system should accomplish one or more of the following objectives. Convey information about past activities, current status or projections of the Future. Signal important events, opportunities, problems, or warnings. Trigger an action. Confirm an action.

CONCLUSION:

Our goal is to propose a deep learning-based approach to classify pneumonia from chest X-ray images using transfer learning. In this framework, we adopted the transfer learning approach and used the pretrained architectures, VGG16 and ResNet trained on the ImageNet dataset, to extract features. These features were passed to the classifiers of respective models, and the output was collected from individual architectures. Finally, we employed an ensemble model that used pretrained models and outperformed all other models. We observed that performance could be improved further, by increasing dataset size, using a data

augmentation approach, and by using hand-crafted features, in future. Our findings support the notion that deep learning methods can be used to simplify the diagnostic process and improve disease management. While pneumonia diagnoses are commonly confirmed by a single doctor, allowing for the possibility of error, deep learning methods can be regarded as a two-way confirmation system. In this case, the decision support system provides a diagnosis based on chest X-ray images, which can then be confirmed by the attending physician, drastically minimizing both human and computer error. Our results suggest that deep learning methods can be used to improve diagnosis relative to traditional methods, which may improve the quality of treatment. When compared with the previous state-of-the-art methods, our approach can effectively detect the inflammatory region in chest X-ray images.

REFERENCES:

1. Asiri, N.; Hussain, M.; Al Adel, F.; Alzaidi, N. Deep learning based computer-aided diagnosis systems fordiabetic retinopathy: A survey. Artif. Intell. Med -2019

2. Bakator, M.; Radosav, D. Deep Learning and Medical Diagnosis: A Review of Literature. Multimodal Technol. Interact, 2, 47- 2018

3. Brunetti, A.; Carnimeo, L.; Trotta, G.F.; Bevilacqua, V. Computer-assisted frameworks for classification of liver, breast and blood neoplasias via neural networks: A survey based on medical images. Neurocomputing , 335, 274–298 - 2019

4. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; van der Laak, J.A.W.M.; Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. Med. Image Anal, 42, 60–88. [PubMed]- 2017

5. Liu, N.; Wan, L.; Zhang, Y.; Zhou, T.; Huo, H.; Fang, T. Exploiting Convolutional Neural Networks With Deeply Local Description for Remote Sensing Image Classification. IEEE Access , 6, 11215–11228 - 2018

6. Malukas, U.; Maskeli ¯ unas, R.; Damaševiˇcius, R.; Wo ´zniak, M. Real time path finding for assisted living using deep learning. J. Univers. Comput. Sci., 24, 475–487- 2018

7. Meyer, P.; Noblet, V.; Mazzara, C.; Lallement, A. Survey on deep learning for radiotherapy. Comput. Biol. Med. 2018, 98, 126–146. Appl. Sci, 10, 559 15 of 17- 2020

8. Shickel, B.; Tighe, P.J.; Bihorac, A.; Rashidi, P. Deep EHR: A survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. IEEE J. Biomed. Health Inform., 22, 1589–1604 - 2018

9. Zhang, X.; Yao, L.; Wang, X.; Monaghan, J.; McAlpine, D. A Survey on Deep Learning based Brain Computer Interface: Recent Advances and New Frontiers. arXiv, arXiv:1905.04149 -2018

10. Zhou, T.; Thung, K.; Zhu, X.; Shen, D. Effective feature learning and fusion of multimodality data using stage-wise deep neural network for dementia diagnosis. Hum. Brain Mapp, 40, 1001–1016 - 2018