upGrad

LOVELY PROFESSIONAL UNIVERSITY

Transforming Education Transforming India

Disease Diagnosis using Medical Imaging and Deep Learning

NAME : KASARLA VISHWAS

REG NO : 12001688

SECTION : K20UG

COURSE NAME : CSE 441 (INDUSTRY INTERNSHIP PROJECT)

MENTOR NAME : Devang

INTRODUCTION

- Medical imaging and deep learning have joined forces, promising to transform how we diagnose and treat diseases. In the past, doctors mainly relied on their judgment to interpret medical images, which could be inconsistent. But now, with deep learning, computers can analyze these images quickly and accurately.
- This breakthrough has caught the attention of many because it offers a way to catch diseases early and improve how we care for people's health. This study aims to explore how these technologies have evolved, how they're changing disease diagnosis, and what challenges and benefits they bring.
- By looking at recent progress and real-life examples, we hope to show how deep learning is changing healthcare for the better.

upGrad

LOVELY PROFESSIONAL UNIVERSITY

Transforming Education Transforming India

PROBLEM STATEMENT

- The rapid evolution of medical imaging techniques, coupled with advancements in deep learning algorithms, presents both opportunities and challenges in disease diagnosis. While deep learning, particularly Convolutional Neural Networks (CNNs), has shown promise in automating the analysis of medical images, there remains a need for a comprehensive understanding of its impact on disease detection and diagnosis.
- Furthermore, the practical implementation of deep learning models in real-world healthcare settings requires addressing various challenges, including data standardization, model optimization, and integration with existing healthcare infrastructure.
- Therefore, the overarching problem addressed by this research is to investigate the effectiveness, limitations, and practical implications of deep learning for medical imaging-based disease diagnosis, with the aim of improving healthcare outcomes and patient care.

DATASET DESCRIPTION

We used a big collection of medical pictures for our project. These pictures come from many places like hospitals and research centers. They show different diseases like COVID-19, pneumonia, malaria, and brain tumors. We have lots of pictures, covering various types of medical scans like X-rays and MRIs.

Each picture is like a piece of information for the computer to learn from. We made sure these pictures were ready for the computer to understand by doing some preparation work. This included making sure all the pictures were the same size and adjusting the colors, so they were easier for the computer to work with. We also made copies of the pictures and changed them slightly to make the computer learn better.

upGrad

LOVELY PROFESSIONAL UNIVERSITY

Transforming Education Transforming India

EXPLORATORY DATA ANALYTICS (EDA)

Data Collection/Extraction:

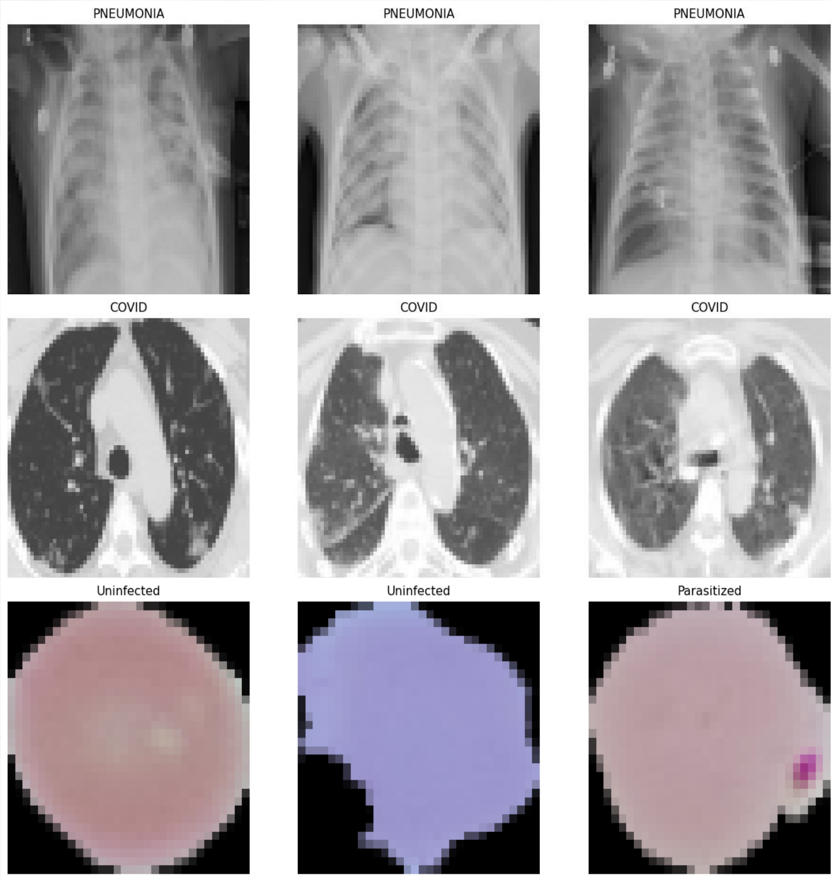
- Sourced medical imaging data from established health databases.
- Compiled datasets of X-rays and MRIs for various diseases.

Data Analysis:

- Univariate Analysis: Examined single variables to understand their distribution and statistics.
- Bivariate Analysis: Explored relationships between two variables to find correlation

Data Predictions:

- Applied deep learning models to predict disease presence from medical images.
- Evaluated model performance using accuracy metrics and validation techniques.



FEATURE ENGINEERING

Features were carefully selected or engineered based on the unique characteristics of medical imaging data.

To improve the performance of our model, we meticulously selected and engineered features from our medical imaging dataset. Initially, we subjected the images to preprocessing steps like normalization, scaling, and augmentation.

These techniques standardized the images and reduced variability, making them more suitable for analysis. Then, we employed advanced methods such as edge detection, texture analysis, and region segmentation to extract features directly from the images. This approach allowed us to identify relevant patterns and structures indicative of various diseases.

Additionally, we incorporated metadata such as patient demographics and clinical history to provide context to the images. Furthermore, dimensionality reduction techniques like Principal Component Analysis (PCA) were applied to handle high-dimensional data and reduce computational complexity. Insights from medical experts guided the selection and engineering of features, ensuring they were clinically meaningful and aligned with our objectives.

MODEL ARCHITECTURE

DenseNet121, our chosen model, has 121 layers and uses dense connectivity, where each layer is connected to every other layer.

This setup helps in better feature reuse and mitigates the vanishing gradient problem.

The model consists of convolutional and pooling layers for extracting features from input data.

It includes bottleneck layers to compress feature maps and skip connections to facilitate gradient flow.

DenseNet121 is efficient and effective for tasks like disease detection in medical images.

TRAINING PROCESS

The training process involved splitting the dataset into training, validation, and test sets. The validation set helped prevent overfitting and fine-tuned hyperparameters, while the training set updated model parameters through backpropagation. Model performance was assessed using the held-out test set.

To optimize the model, the categorical cross-entropy loss function was minimized using classic optimization techniques like Adam or stochastic gradient descent (SGD) with momentum. Regularization techniques such as weight decay and dropout were applied to enhance model generalization and prevent overfitting. Learning rate scheduling was utilized to maximize training efficiency.

Various evaluation criteria, including accuracy, precision, recall, F1-score, and ROC curves, were used to assess the model's performance, providing insights into its classification accuracy and reliability on unseen data.

MODEL EVALUATION

In diagnosing pneumonia, both ML classifiers and DL models are utilized, with considerations given to labeled datasets and feature extraction methods. Evaluation metrics include accuracy, sensitivity, and specificity.

In lung disease diagnosis, DL models and ML approaches are used, emphasizing data preprocessing and feature selection. Evaluation metrics include AUC, F1 score, and confusion matrix analysis.

Malaria detection involves ML classifiers and convolutional networks, analyzing blood smear images for parasite detection. Evaluation metrics include sensitivity, specificity, and positive predictive value (PPV).

For detecting multiple diseases, CNN architectures and ML classifiers are utilized on high-dimensional datasets and computer-aided diagnosis (CAD) systems. Evaluation metrics consist of accuracy, positive predictive value (PPV), and negative predictive value (NPV).

RESULTS AND ANALYSIS

The results from testing the model showed that it performed well in diagnosing different diseases like pneumonia, brain disorders, lung diseases, malaria, and multiple diseases at once.

For each disease, the model could accurately identify whether it was present or not, showing good accuracy in its predictions.

It also did a good job of catching specific cases, like detecting pneumonia or brain diseases in MRI scans.

However, there's still room for improvement, like tweaking some settings to make the model even better and using more diverse data for training to cover a wider range of cases.

Overall, the model shows promise for helping doctors diagnose diseases more accurately, but there's still work to be done to make it even more effective

DEPLOYMENT

To deploy my machine learning project, I first organize all the necessary parts, like UI code and ML models.

Then, I sign up for Render.com and connect my GitHub repo. Next, I set up the deployment by specifying the environment and configurations.

After starting the deployment, I check for any issues and fix them if needed.

Once deployed, I thoroughly test the app and make continuous updates using Render.com's tools. Regular monitoring ensures the app stays efficient and user-friendly.

CHALLENGES AND LIMITATIONS

1. **Data Quality and Variability**: Ensuring the accuracy and reliability of machine learning models in real-world scenarios can be challenging due to variations in data quality, distribution, and dynamics.

2. **Integration Complexity**: Integrating the models into existing systems or platforms may require substantial effort to ensure compatibility and seamless operation, adding complexity to the deployment process.

3. **Scalability**: Scaling the project to handle large volumes of data or user requests while maintaining performance and efficiency poses a significant challenge, requiring careful consideration of infrastructure requirements and scalable solutions.

4. **Data Privacy and Security**: The project may face limitations related to data privacy and security, especially when dealing with sensitive information, requiring compliance with regulatory requirements and implementation of robust security measures.

5. **Maintenance and Updates**: Ongoing maintenance and updates are necessary to address bugs, improve model performance, and adapt to changing user needs or environmental conditions, requiring dedicated resources and processes for effective deployment and management.

FUTURE WORK

1. **Adding More Data Insights**: We can look at more details from different sources to improve our model, like understanding text better or using multiple types of data together.

2. **Making the Model Better**: We can try using smarter models that can understand more complicated patterns in the data, which could help our predictions be more accurate.

3. **Learning from Existing Models**: We can use models that already know a lot about similar things and teach them more about our specific problem, which might make them better at predicting.

4. **Combining Different Models**: We can put together predictions from several models to make a final decision, kind of like asking different experts for their opinions and then combining them.

5. **Understanding Why the Model Decides**: We can make sure the model's decisions are easier to understand, like highlighting which parts of the data are most important for its decisions, especially in areas like healthcare or money where people need to trust the model's decisions.

CONCLUSION

This study emphasizes the transformative impact of deep learning on medical imaging and diagnosis, offering insights into its status and prospects.

Through a case study utilizing DenseNet121, it illustrates how deep learning can bolster diagnostic capabilities for illnesses like COVID-19, pneumonia, malaria, and brain tumors, leading to improved patient outcomes.

The findings advocate for continued exploration, collaboration, and innovation to realize AI's full potential in healthcare, while also addressing ethical and regulatory considerations for responsible implementation in clinical practice.

upGrad

LOVELY PROFESSIONAL UNIVERSITY

Transforming Education Transforming India

REFERENCES

Y. Chen, Q. Liu and D. Guo, "Emerging coronaviruses: Genome structure replication and pathogenesis", J. Med. Virol., vol. 92, no. 4, pp. 418-423, 2020. Show in Context CrossRef Google Scholar

2. Q. Li et al., "Early Transmission Dynamics in Wuhan China of Novel Coronavirus–Infected Pneumonia", N. Engl. J. Med., vol. 382, no. 13, pp. 1199-1207, 2020. Show in Context CrossRef Google Scholar

3. A. Bustamam, E. D. Ulul, H. F. A. Hura and T. Siswantining, "Implementation of hierarchical clustering using k-mer sparse matrix to analyze MERS-CoV genetic relationship", AIP Conf. Proc, vol. 1862, no. Iscpms, pp. 1-7, 2017. Show in Context CrossRef Google Scholar

4. Coronavirus Disease (COVID-19) Outbreak Situation, [online] Available: https://www.who.int/emergencies/disease/novel-coronavirus-2019. Show in Context Google Scholar

THANK YOU