KCES'S INSTITUTE OF MANAGEMENT AND RESEARCH (AUTONOMOUS), JALGAON

MCA-RP-638 : Minor Project (Research / S/W Development)

Student Names:

1) Jayesh Prashantbhai Attarde (R.No. 03)
2) Harshal Sunil Bhangale (R. No.10)

Project Title :

Pneumonia Detection from Chest X-ray Using Convolutional Neural Network

Under the guidance of

Project Guide : Prof. Dr. Deepali Y. Kirange

❖ Input (Description):

The system primarily accepts Chest X-ray images uploaded by users through a web interface. These images can belong to one of three categories: NORMAL, PNEUMONIA, or INVALID (e.g., non-chest images or unclear scans). The main input flow includes:

- **User Uploads**: Users can upload chest X-rays through a browser-based interface.

- **Image Requirements**: Images should be clear, correctly oriented, and in formats such as .jpg, .jpeg, or .png.

- **User Data**: During signup or login, user information like Name, Email, username, Phone Number and password is also collected and validated.

- **Dataset**: For model training and testing, labeled data is taken from a dataset folder, which is structured into train, test, and validation subfolders.

Before feeding into the model, uploaded images are resized, normalized, and converted to grayscale or RGB if required. This preprocessing ensures consistency for CNN-based image analysis.

❖ Output (Description)

The main output of the system is a prediction result that classifies the uploaded image as either NORMAL, PNEUMONIA, or INVALID. This output is generated using a trained Convolutional Neural Network (CNN) model.

<u>Output details include:</u>

- **Prediction Label**: The model gives one of three labels based on the X-ray:
  ➤ NORMAL – Healthy lungs
  ➤ PNEUMONIA – Infected lungs
  ➤ INVALID – Non-X-ray or unclear image

- **Confidence Score**: A probability (e.g., 93.6%) showing how confident the model is in its prediction.
- **Visual Feedback**: The result is displayed on the user’s dashboard in a clean, easy-to-read format.
- **Image Storage**: Uploaded images and their results are saved in the system for historical review.

This output enables early detection and improves decision-making, especially where medical consultation is not immediately available.

❖ Existing Work (Survey / Review)

A. The Need for Pneumonia Detection Using AI

Pneumonia is a potentially life-threatening lung infection that causes inflammation in the air sacs (alveoli), often leading to fluid accumulation, breathing difficulty, and decreased oxygen intake. It can be caused by bacteria, viruses, or fungi.

Early and accurate diagnosis of pneumonia is essential for initiating timely treatment and preventing complications. Traditionally, this diagnosis relies on physical symptoms, patient history, and chest X-ray (CXR) interpretation by radiologists.

However, there are several real-world challenges in this process:

- **Radiologist Shortage**: Many rural or underdeveloped areas lack access to trained radiologists.
- **Time Constraints**: In busy hospitals, radiologists face large workloads, delaying diagnosis.
- **Human Error**: Fatigue or misinterpretation can lead to false positives/negatives.

This creates a strong case for Artificial Intelligence (AI)-based solutions, especially deep learning models that can analyze chest X-ray images automatically and flag potential pneumonia cases. Such tools can act as decision support systems, offering second opinions or even working autonomously in areas without human experts.

B. Traditional Image Processing vs AI-Based Diagnosis

1) Traditional/Manual Techniques:

Before AI, image classification tasks were approached using traditional image processing techniques such as:

- **Edge Detection**: Identifying outlines of anatomical structures using Sobel or Canny filters.
- **Thresholding**: Segmenting images based on grayscale intensity.
- **Texture Analysis**: Using statistical features (GLCM, LBP) to detect irregularities.
- **Feature Engineering**: Manually selecting features like pixel intensity, histogram patterns, etc.

While these methods worked in controlled environments, they had several limitations:

- Poor generalization to diverse or noisy datasets.
- Required expert intervention for feature selection.
- Lacked scalability and flexibility.

---

2) AI & Deep Learning Solutions :

The emergence of Convolutional Neural Networks (CNNs) revolutionized image classification. CNNs automatically learn features from raw pixel data through hierarchical layers. In the medical field, CNNs have demonstrated human-level or superior accuracy in interpreting diagnostic images.

Key advantages of CNN-based models:

- No manual feature extraction required.
- Generalize well to complex image variations.
- Improve continuously with more data (transfer learning, fine-tuning).

C. Related Research and Existing Tools :

Several landmark studies and open-source tools have laid the foundation for AI-based pneumonia detection from chest X-rays:

1. CheXNet – Stanford University (2017) :

CheXNet is a deep learning model based on DenseNet-121, trained on ChestX-ray14, a large dataset with over 100,000 frontal-view chest X-rays. It was developed by Stanford researchers and achieved expert-level accuracy in pneumonia detection.

- **Architecture**: 121-layer DenseNet
- **Input**: Frontal-view X-rays
- **Output**: Probabilities for 14 diseases, including pneumonia
- **Accuracy**: Outperformed practicing radiologists in ROC-AUC
- **Limitation**: Black-box nature, requires significant computational power.

CheXNet popularized the idea that AI can match or surpass human radiologists, bringing credibility to deep learning in healthcare.

2. Kermany et al. Pediatric Pneumonia Dataset (2018) :

Kermany and colleagues introduced a public dataset of pediatric chest X-rays containing labeled samples for Normal, Pneumonia (Viral), and Pneumonia (Bacterial). It includes:

- 5,863 labeled images
- High-quality annotations
- Widely used for benchmarking pneumonia classification models

Their work used a relatively simple CNN but still achieved high accuracy due to clean and curated data. It remains one of the most cited datasets in medical AI.

3. NIH ChestX-ray14 Dataset :

Released by the National Institutes of Health (NIH), this dataset includes:

- Over 100,000 chest X-rays
- Annotations for 14 disease types
- Tools for bounding box generation

Although powerful, the ChestX-ray14 dataset has some labeling issues, as many annotations were automatically extracted from radiology reports (prone to errors).

4. COVID-Net (2020) :

Developed in response to the COVID-19 pandemic, COVID-Net is a CNN model trained on X-rays and CT scans for detecting COVID-19 pneumonia.

- Open-source project
- Focused on COVID detection but validated pneumonia patterns
- Demonstrates adaptability of CNNs for different lung diseases

COVID-Net helped extend pneumonia detection research into pandemic scenarios and proved the reusability of deep learning architectures across diseases.

Though we rely solely on Paul Mooney’s dataset, our work is inspired by these larger initiatives, applying deep learning methods in a more accessible and deployable format.

D. Limitations of Existing Work

Despite their contributions, current systems face several practical and technical challenges that limit their real-world deployment.

1. **Model-Centric Focus:**

Many research papers focus purely on building high-performing CNN models but ignore the end-user experience, such as how a doctor, technician, or patient will interact with the system. There is often no web interface or input validation mechanism.

2. **Lack of Real-World Integration:**

Even high-accuracy models like CheXNet are rarely deployed with:

- Web-based frontends
- Secure user authentication
- Database integration
- User role management (admin, doctor, patient)

This makes them unsuitable for clinical use without significant additional development.

3. **High Hardware Requirements :**

Training and even inference of some deep models require:

- Powerful GPUs (e.g., NVIDIA Tesla, RTX series)
- High RAM and disk storage
- Cloud compute credits (for Google Colab, AWS, etc.)

These constraints make deployment in rural clinics or small hospitals impractical.

4. **No Invalid Input Detection :**

Most existing systems assume the uploaded image is valid, i.e., a chest X-ray. In practice, users may accidentally upload unrelated images. Without proper validation, the model may make wrong predictions, leading to dangerous outcomes.

E. How Our Project Bridges These Gaps :

Our project proposes a practical, integrated, and user-friendly system that addresses the above limitations:

- Uses a pre-trained CNN for pneumonia detection.
- Implements a complete Django web app for user interaction.
- Provides user registration, login, and admin dashboard.
- Connects to MongoDB Atlas for scalable and secure data storage.
- Supports image upload with format and content validation.
- Runs on moderate hardware (can be deployed locally or on the cloud).
- Saves prediction results for review and audit.
- Includes an INVALID class to handle wrong inputs safely.

By combining machine learning, web development, and cloud database integration, the project is closer to a real-world deployable solution than many academic prototypes.

❖ Proposed Work

The goal is to create a web-based pneumonia detection platform that combines deep learning with a user-friendly interface.

Key Objectives:

- Upload chest X-ray images via a browser.
- A trained CNN model to predict if pneumonia is present.
- Classify images into NORMAL, PNEUMONIA, or INVALID categories.
- Securely store user data and uploaded images.
- Provide separate dashboards for users and administrators.
- Enable real-time use in hospitals, clinics, and rural settings.

Innovations:

- Integration of AI and web framework (Django).
- MongoDB for cloud-based, flexible data storage.
- Clean and modern UI with Bootstrap for smooth navigation.
- Invalid input handling for safety and accuracy.

This work not only demonstrates technical capability but also contributes to real-world healthcare applications.

❖ Methodology of Work

Step 1: **Dataset Preparation :**

- Dataset is collected from Kaggle and includes:
  X-rays are labelled as NORMAL, PNEUMONIA, INVALID**.**
- Images are resized to 150x150 pixels, normalized, and augmented for better generalization.

Step 2: **Model Building :**

- Built using Keras and TensorFlow.
- CNN architecture includes layers:
  - Conv2D for feature extraction
  - MaxPooling2D for downsampling
  - Dropout for regularization
  - Dense for classification
- Trained using categorical cross-entropy loss and Adam optimizer.

Step 3: **Model Saving**

- Final trained model is saved as pneumonia_model.keras and stored in the ml_model/ folder.

Step 4: **Web Development (Django)**

- Django backend handles routing, form submissions, and predictions.
- Templates provide frontend views for user and admin.
- views.py manages upload, prediction, and dashboard logic.

Step 5: **Database Integration**

- User details and image metadata are stored in MongoDB Atlas.
- Secure connection managed via .env file.

❖ Features of the System

- Upload chest X-ray images securely.
- Classify images into NORMAL / PNEUMONIA / INVALID.
- Get real-time results with model confidence score.
- User login/signup with form validation.
- Admin dashboard to manage users and image uploads.
- Uses MongoDB Atlas for flexible, cloud-based data storage.
- Saves image history for each user session.
- Clean, responsive web design.

❖ Purpose / Use of the Project

This system serves as an AI-based diagnostic assistant designed to detect pneumonia from chest X-rays. Its main purposes are:

- **Assist Radiologists:** Reduce workload and speed up diagnosis.
- **Use in Clinics/Hospitals:** Useful in busy or under-staffed healthcare centers.
- **Rural Deployment**: Can be used in telemedicine setups for remote screening.
- **Educational Tool:** Demonstrates the use of CNNs in healthcare.
- **Secure Medical App:** Data is stored safely with login and admin access.

The project aims to be scalable, fast, and reliable, making it suitable for real-world healthcare applications and academic exploration.

❖ REFERENCES

1. CheXNet (Stanford University)

Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., & Ng, A. Y. (2017).
CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning.

Link : https://arxiv.org/abs/1711.05225

2. Kermany et al. – Pediatric Chest X-ray Dataset

Kermany, D. S., Zhang, K., & Goldbaum, M. (2018). Identifying medical diagnoses and treatable diseases by image-based deep learning.

Link : https://www.cell.com/fulltext/S0092-8674(18)30154-5

3. NIH ChestX-ray14 Dataset

Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., & Summers, R. M. (2017).
ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases.

Link : https://arxiv.org/abs/1705.02315

4. COVID-Net

Wang, L., Lin, Z. Q., & Wong, A. (2020).
COVID-Net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest radiography images. Scientific Reports, 10(1), 19549.
Link : https://www.nature.com/articles/s41598-020-76550-z

5. Paul Mooney's Chest X-ray Dataset

Mooney, P. (2018). Chest X-Ray Images (Pneumonia). Kaggle.
Link : https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia