FINAL YEAR PROJECT REPORT

PEDIATRIC CHEST DISEASE PREDICTION

In fulfillment of the requirement
For degree of
BS (COMPUTER SCIENCES)

By

RAAHIMA IRFAN	**70110 (BSCS)**
MUSADDIQ ALI SADIQ	**70145 (BSCS)**

SUPERVISED

BY
MS.TOOBA MEHTAB

BAHRIA UNIVERSITY (KARACHI CAMPUS)

SPRING-2024

DECLARATION

We hereby declare that this project report is based on our original work except for citations and quotations which have been duly acknowledged. We also declare that it has not been previously and concurrently submitted for any other degree or award at Bahria University or other institutions.

Signature : ____________________

Name : MUSADDIQ ALI SADIQ

Reg No. : 70145

Signature : ____________________

Name : RAAHIMA IRFAN

Reg No. : 70110

Date : 14/06/2024

APPROVAL FOR SUBMISSION

We certify that this project report entitled **"PEDIATRIC CHEST DISEASE PREDICTION"** was prepared by **MUSADDIQ ALI SADIQ** and **RAAHIMA IRFAN** has met the required standard for submission in partial fulfilment of the requirements for the award of Bachelor of Science in Bachelors' of Science in Computer Science at Bahria University, Karachi.

Approved by,

Signature : ____________________

Supervisor : Ms. TOOBA MEHTAB

Date : 14/06/24

ACKNOWLEDGEMENTS

We would like to thank everyone who had contributed to the successful completion of this project. We would like to express my gratitude to my project supervisor, Ms. TOOBA MEHTAB for her invaluable advice, guidance and her enormous patience throughout the development of the research.

In addition, we would also like to express our gratitude to our loving parents and friends who have helped and given us encouragement.

PEDIATRIC CHEST DISEASE PREDICTION

ABSTRACT

This study addresses the classification of pediatric chest X-ray images into three diagnostic categories: Infiltration, Effusion, and Atelectasis. Utilizing the NIH Chest X-ray dataset filtered for patients aged 12 or younger, we employ convolutional neural networks (CNNs) for feature extraction and rigorously compare various machine learning and deep learning models, including Random Forest, Logistic Regression, SVM, Decision Tree, and DenseNet121. A key contribution of our research is the use of a pre-trained VGG16 model for feature extraction, followed by the application of a Random Forest classifier, which achieved a high classification accuracy. Our experimental methodology involves addressing class imbalance through dataset balancing techniques and evaluating models on differently sized datasets. Results indicate that larger datasets substantially enhance model performance, with DenseNet121 attaining the highest overall accuracy and AUC scores. Notably, our VGG16-augmented Random Forest model also demonstrated robust performance, highlighting the effectiveness of our feature extraction and classification pipeline. The findings of this study underscore the necessity of adequate training data and validate the superiority of deep learning models in handling complex medical imaging tasks. This research underscores the potential of advanced machine learning techniques in improving pediatric chest X-ray classification, thus enhancing diagnostic accuracy and supporting clinical decision-making. Future research directions will explore data augmentation, hyperparameter optimization, ensemble methods, and strategies to improve model interpretability and generalizability.

TABLE OF CONTENTS