

CONTENTS

PREFACE	iv
CONTENTS	v
LIST OF FIGURES	viii
LIST OF TABLES	xi
CHAPTER 1 INTRODUCTION	1
1.1 Research Background	1
1.2 Research Problem	3
1.3 Research Scope	4
1.4 Research Objective	5
1.5 Research Benefits	5
CHAPTER 2 LITERATURE REVIEW	6
2.1 Deep Learning	6
2.2 Data Balancing.....	8
CHAPTER 3 THEORETICAL BASIS	14
3.1 Image Classification	14
3.2 Deep Learning Methods	14
3.2.1 DenseNet-121	15
3.3 Dataset	16
3.4 Data Balancing Techniques	17
3.4.1 SMOTE	17
3.4.2 Data Augmentation	19
3.4.3 Random Oversampling.....	19
3.5 Evaluation Metrics	20
3.5.1 Learning Curve.....	20
3.5.2 Loss Function	21
3.5.3 Confusion Matrix	22

3.6	Renal Disease.....	23
CHAPTER 4 RESEARCH METHODOLOGY		26
4.1	Research Description	26
4.2	Research Phases	27
4.2.1	Data Collection.....	27
4.2.2	Data Preprocessing.....	28
4.2.3	Imbalance Handling	28
4.2.4	Data Splitting	29
4.2.5	Model Development and Training	29
4.2.6	Model Evaluation	31
CHAPTER 5 IMPLEMENTATION		33
5.1	Data Collection	33
5.2	Data Preprocessing and Exploratory	35
5.3	Imbalance Data Handling	41
5.3.1	SMOTE	41
5.3.2	Data Augmentation	42
5.3.3	Random Oversampling.....	44
5.4	Data Splitting	45
5.5	Model Development and Training	46
5.6	Model Evaluation.....	50
5.6.1	Learning Curve.....	50
5.6.2	Confusion Matrix and Classification	51
CHAPTER 6 RESULTS AND DISCUSSION		54
6.1	Data Balancing Results.....	54
6.2	Model Performance Evaluation Results	55
6.2.1	Without data balancing	55
6.2.2	SMOTE	56
6.2.3	Data Augmentation	57
6.2.4	Random Oversampling.....	59



CHAPTER 7 CONCLUSION AND FUTURE WORKS	61
7.1 Conclusion	61
7.2 Future Works.....	62
REFERENCES.....	64

LIST OF FIGURES

Figure 1.1 Loss curve and learning curve of preliminary experiment	3
Figure 1.2 Classification report of preliminary experiment.....	4
Figure 3.1 Architecture of DenseNet-121 (Tiwari et al., 2023).....	16
Figure 3.2 Visualization of how SMOTE works (Singh, 2023). Link to article... 18	
Figure 3.3 Example result of data augmentation (Wang & Perez, 2017)	19
Figure 3.4 Example of learning curve showing model performance over epochs (Ibrahim, 2023). Link to article.....	20
Figure 4.1 Flowchart of the research process.....	26
Figure 4.2 Research phases	27
Figure 4.3 Sample CT scan images from the dataset used in the research	28
Figure 5.1 Python script to get the image path to each cases and list all the images	34
Figure 5.2 Python script to create an empty list and insert the train data in (img_path, label) format	34
Figure 5.3 Python script to create a pandas DataFrame from the data in the list..	34
Figure 5.4 First five rows of the created pandas DataFrame	35
Figure 5.5 Python script and output of checking the unique labels	35
Figure 5.6 Python script and output for checking 'train_data' dimension.....	35
Figure 5.7 Python script for visualizing the class distribution.....	36
Figure 5.8 Visualization of class distribution of the dataset	36
Figure 5.9 Python script for getting the image preview of the dataset	37
Figure 5.10 Image preview of the dataset (the output of the previous Python script)	38
Figure 5.11 Python script for preprocessing the data by converting the data into NumPy arrays.....	39
Figure 5.12 Python script to save preprocessed data into .npy and .pkl files for efficient reuse	40
Figure 5.13 Python script to load preprocessed data and verify the shapes of train_data and train_labels	40

Figure 5.14 Code snippet demonstrating data balancing using SMOTE	41
Figure 5.15 Python script to define the data augmentation settings	42
Figure 5.16 Python script for imbalance handling using data augmentation (geometric transformations)	44
Figure 5.17 Python script to resample data with random oversampling with equal distribution	45
Figure 5.18 Python script to reshape the resampled data.....	45
Figure 5.19 Python script to implement data splitting on the dataset	47
Figure 5.20 Python script for defining the function for DenseNet-121 model	48
Figure 5.21 Python script for model implementation	49
Figure 5.22 Code implementation to visualize learning curve	51
Figure 5.23 Code implementation to predict the testing set using the trained model	52
Figure 5.24 Python script to print display confusion matrix.....	52
Figure 5.25 Python script to display classification report.....	53
Figure 6.1 Data distribution bar chart before data balancing.....	54
Figure 6.2 Data distribution bar chart after balancing	55
Figure 6.3 Learning curve for model's performance without imbalance handling	56
Figure 6.4 Classification report of model's prediction without imbalance handling	56
Figure 6.5 Learning curve for model performance with SMOTE to handle data imbalance	57
Figure 6.6 Classification report of model's prediction by using SMOTE to handle data imbalance.....	57
Figure 6.7 Learning curve for model performance by using data augmentation (geometric transformations) to handle data imbalance	58
Figure 6.8 Classification report of model's prediction by using data augmentation to handle data imbalance	58
Figure 6.9 Learning curve for model performance with Random Oversampling to handle data imbalance	59



Figure 6.10 Classification report of model's prediction by using random oversampling to handle data imbalance 60



LIST OF TABLES

Table 2.1 Works Related to Deep Learning Methods Implemented on Medical Image Dataset.....	7
Table 2.2 Works Related to Data Balancing Techniques Implemented on Non-Medical Image Dataset.....	8
Table 2.3 Works Related to Data Balancing Techniques Implemented on Medical Image Dataset.....	11
Table 4.1 Experimentation scheme	30
Table 4.2 Model flow architecture of DenseNet-121 (Huang et al., 2018)	31
Table 7.1 Summary of model's performance	62