



Ministry of Higher Education and  
Scientific Research  
University of Diyala  
College of Science  
Department of Computer Science



# *Thalassemia Disease Classification Based On Machine Learning Techniques*

A Research

Submitted to the Department of Computer Science\  
College of Sciences\ University of Diyala in a Partial  
Fulfillment of the Requirements for the Degree of  
Master in Computer Science

*By*

*Muna Qais Mohammed*

Supervised By

*Assist. Prof. Dr. Jamal Mustafa  
Al-Tuwaijari*

2021A.D

1443 A.H

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

﴿ تَرْفَعُ دَرَجَاتٍ مَن نَّشَاءُ وَفَوْقَ كُلِّ ذِي عِلْمٍ عَلِيمٌ ﴾

صَدَقَ اللَّهُ الْعَظِيمُ

سورة يوسف ( آية 76 )

## *Examination Committee Certification*

*We certify that we have read the thesis entitled “Thalassemia Disease Classification And Detection Based On Artificial Intelligence Technique” and an examination committee, examined the student “Muna Qais Mohammed” in the thesis content and our opinion it is adequate as fulfill the requirement for Degree of Master in Computer Science at the Computer Science Department, University of Diyala.*

Signature:

Name: **Prof.Dr. Ziad Tariq Mustafa** (Chairman)

Date: / /2021

Signature:

Name: **Assist.Prof.Dr. Khalid Shaker Jasim** (Member)

Date: / /2021

Signature:

Name: **Assist.Prof.Dr. Bashar Talib Hameed** (Member)

Date: / / 2021

Signature:

Name: **Assist. Prof. Dr. Jamal Mustafa Al-Tuwaijari** (Member/Supervisor)

Date: / / 2021

Approved by the **Dean** of College of Science, University of Diyala.

(The Dean)

Signature:

Name: **Prof. Dr. Tahseen H. Mubarak**

Date: / / 2021

## (Supervisor Certification)

*I certify that this research entitled “Thalassemia Disease Classification And Detection Based On Artificial Intelligence Technique” was prepared by “Muna Qais Mohammed” Under my supervision at the Department of Computer Science\ College of Sciences\ University of Diyala in a Partial Fulfillment of the Requirements for the Degree of Master in Computer Science.*

(Supervisor)

Name: **Assist. Prof. Dr. Jamal Mustafa Al-Tuwaijari**

Signature:

Date: / /2021

*Approved by Department of Computer Science\ College of Sciences\ University of Diyala*

Signature:

Name: **Assist. Prof. Dr. Bashar Talib Hameed**

Date: / /2021

*Head of Department of Computer Science\ College of Sciences\ University of Diyala*

## **(Linguistic Certification)**

*I certify that the thesis entitled “Thalassemia Disease Classification And Detection Based On Artificial Intelligence Technique” was prepared by “Muna Qais Mohammed” and was reviewed linguistically. Its language was amended to meet the style of the english language.*

Signature:

Name:

Date: / /2021

## **(Scientific Amendment)**

*I certify that the thesis entitled “Thalassemia Disease Classification And Detection Based On Artificial Intelligence Technique” was prepared by “Muna Qais Mohammed” has been evaluated scientifically ; therefore, it is suitable for debate by examining committee.*

Signature:

Name:

Date: / / 2021

## ACKNOWLEDGMENT

Initially and foremost I deeply thank God, Lord of the worlds, for all the grace and helping to bring this thesis to its end.

I would like to grateful my supervisor Dr Jamal Mustafa AL-Tuwaijari for his support and encouragement. I am extremely thankful for his effort, completing and revising the thesis. My sincere thanks and gratitude is for the Council of the College of Science at Diyala University and the staff of Computer Science Department for all the supports to perform this study and for giving me the opportunity to have the M.Sc.

I would like to express my gratitude to my father and mother, my daughter , my brothers and their families and my sister who had unlimited support and patience.

There are not enough words to thank my dear husband *Bassam* for his logistic support and belief and encouragement throughout my studies. I pass many Thanks to my colleagues *Hayder Al-Bawe and Hafssa* .

Finally, I apologize for all those who not mentioned here and I thank them a lot.



*muna q. mohammed*

## ***DEDICATION***

*To the flower of life and its light and the most precious person in my life, my tender mother.*

*To whom I proudly carry your name, teach me how to make success and instill confidence in myself my dear father, may God extend your life.*

*To those who supported me in adversity and the source of my happiness, my companion to my path and my love, my dear husband.*

*To whom their love and blood flow in my veins, and I lived with them the most beautiful moments, my brothers and sisters.*

*To the hony of my love and my soul, my dear my daughter.*

*To the candles that illuminate the path of knowledge in my path, my distinguished teachers.*

*To everyone who loved me and supported me in my scientific and practical life.*



*muna q. mohammed*



# ABSTRACT

Thalassemia is considered one of the most common genetic blood disorders that has received excessive attention in the medical research fields worldwide. It cannot be cured, but an early detection and classification using screening process is the best way to prevent the disease. If early classification is done, patients can get the right treatment. It helps them increase their life expectancy and reduce the risk of thalassemia to the next generation.

In this thesis, efficient thalassemia classification system have been design to increase the accuracy and decrease the error rate in the diagnosis process. This system based on two proposed approaches for classifying thalassemia disease. The first proposed approach is based on four machine learning technique which include artificial neural network (ANN), decision tree (DT), k-nearest neighbor (KNN) and logistic regression (LR). This approach consist of two main stages: pre-processing and classification of thalassemia disease. The second proposed approach based on two deep learning techniques, these are consist of convolutional neural network (CNN) and deep neural network (DNN).

The proposed system has been tested by using two thalassemia dataset. The first type of dataset contain 391 sample with nine features and split into 30% for testing and 70% for training and the second type of dataset contain 7108 image with nine type of erythrocytes and split into 20% for testing and 80% for training. The comparison results show that the proposed system has efficient diagnosis performance and can be used as a promising tool for thalassemia disease diagnosis compared to another existing method where the accuracy rate of first proposed approach for first dataset using ANN is 99%, DT is 99%, LR is 99% and KNN is 98%. The result of second proposed approach for the second dataset using deep learning technique has an average accuracy of 99% for CNN and 83% for DNN.

# ***List of Contents***

<b><i>Subject</i></b>	<b><i>Page No.</i></b>
<i>List of Contents</i>	<i>I</i>
<i>List of Abbreviations</i>	<i>VI</i>
<i>List of Tables</i>	<i>VIII</i>
<i>List of Figures</i>	<i>IX</i>
<i>List of Algorithms</i>	<i>IX</i>
<b><i>Chapter One: General Introduction</i></b>	
<i>1.1 Introduction</i>	<i>1</i>
<i>1.2 Related Works</i>	<i>3</i>
<i>1.3 Problem Definition</i>	<i>8</i>
<i>1.4 Aims of the Thesis</i>	<i>9</i>
<i>1.5 Thesis challenges</i>	<i>9</i>
<i>1.6 Thesis Outlines</i>	<i>10</i>
<b><i>Chapter Two: Theoretical Background</i></b>	
<i>2.1 Introduction</i>	<i>12</i>
<i>2.2 Thalassemia Disease</i>	<i>12</i>
<i>2.3 Preprocessing</i>	<i>13</i>
<i>2.3.1 dropping</i>	<i>14</i>
<i>2.3.2 Data Cleaning</i>	<i>14</i>
<i>2.3.3 Data normalization</i>	<i>14</i>

<i>2.4 Classification</i>	<i>15</i>
<i>2.5 Machine Learning techniques</i>	<i>16</i>
<i>2.5.1 Artificial Neural Network (ANN)</i>	<i>18</i>
<i>2.5.1.1 Activation Function</i>	<i>19</i>
<i>2.5.1.2 Back Propagation Neural Networks</i>	<i>20</i>
<i>2.5.2 Decision Tree</i>	<i>22</i>
<i>2.5.3 Logistic Regression</i>	<i>25</i>
<i>2.5.4 K- Nearest Neighbor (KNN)</i>	<i>27</i>
<i>2.6 Medical Image Analysis</i>	<i>28</i>
<i>2.7 Deep Learning</i>	<i>29</i>
<i>2.8 Convolution Neural Network (CNN)</i>	<i>30</i>
<i>2.8.1 Convolution operation</i>	<i>31</i>
<i>A. The Size and Number of Filters</i>	<i>33</i>
<i>B. Stride</i>	<i>34</i>
<i>C. Padding</i>	<i>35</i>
<i>D. The formula of the Convolutional</i>	<i>36</i>
<i>2.8.2 Nonlinearity</i>	<i>36</i>
<i>2.8.3 Pooling Layer</i>	<i>36</i>
<i>2.8.4 Dropout Network</i>	<i>37</i>
<i>2.8.5 Normalization Layer in Neural Networks</i>	<i>38</i>
<i>2.8.6 SoftMax Layer</i>	<i>39</i>
<i>2.9 Loss Function</i>	<i>40</i>

<i>2.10 Stochastic Gradient Descent with Adamax (SGDA)</i>	<i>41</i>
<i>2.11 Multilayer Neural Network (MLNN)</i>	<i>42</i>
<i>2.12 Evaluation Measures</i>	<i>43</i>
<i>A. Accuracy</i>	<i>44</i>
<i>B. Precision</i>	<i>44</i>
<i>C. Recall</i>	<i>44</i>
<i>D. F1-Score</i>	<i>45</i>
<i>E. Calculating the number of Parameters</i>	<i>45</i>
<b><i>Chapter Three: The Proposed System Implementation</i></b>	
<i>3.1 Introduction</i>	<i>48</i>
<i>3.2 The proposed Models</i>	<i>48</i>
<i>Case1: The First Proposed Approach Is Classification with Machine Learning Technique</i>	<i>48</i>
<i>3.3 Pre-processing Stage</i>	<i>49</i>
<i>3.3.1 Dropping</i>	<i>50</i>
<i>3.3.2 Data cleaning</i>	<i>50</i>
<i>3.3.3 Data Normalization</i>	<i>50</i>
<i>3.4 The Proposed Technique Of The First Approach</i>	<i>51</i>
<i>3.4.1 Artificial Neural Networks</i>	<i>51</i>
<i>A. Training Phase</i>	<i>52</i>
<i>B. Testing Phase</i>	<i>52</i>
<i>3.4.2 Decision Tree</i>	<i>54</i>

<i>A. Training Phase</i>	55
<i>B. Testing Phase</i>	55
<i>3.4.3 Logistic regression</i>	56
<i>A. Training Phase</i>	56
<i>B. Testing Phase</i>	56
<i>3.4.4 K-Nearest-Neighbor</i>	57
<i>A. Training Phase</i>	58
<i>B. Testing Phase</i>	58
<i>Case2: The Second Proposed Approach: Classification with Deep Learning</i>	59
<i>3.5 Image Pre-processing Stage</i>	60
<i>3.6 Data visualization</i>	61
<i>3.7 Convolutional Neural Network</i>	62
<i>3.8 Data Preparation for CNN Training</i>	63
<i>3.9 The proposed Network Architecture</i>	64
<i>3.9.1 Input layer</i>	66
<i>3.9.2 Convolution Layer</i>	66
<i>3.9.3 Non-linear Layer (Activation Function)</i>	66
<i>3.9.4 Pooling Layer</i>	66
<i>3.9.5 Normalization Layer</i>	67
<i>3.9.6 Fully Connected Layer</i>	68
<i>3.9.7 Softmax Activation Function</i>	68

<i>3.10 CNN Training</i>	<i>69</i>
<i>3.5.5 Training Option (Training Algorithm)</i>	<i>70</i>
<i>A. Stochastic Gradient Descent with AdaMax Optimizer</i>	<i>70</i>
<i>B. Max Epoch</i>	<i>70</i>
<i>C. Shuffle</i>	<i>71</i>
<i>D. Loss Function</i>	<i>71</i>
<i>3.6 CNN Testing</i>	<i>71</i>
<i>3.7 Deep Neural Network</i>	<i>72</i>
<i>A. Training Phase</i>	<i>73</i>
<i>B. Testing Phase</i>	<i>73</i>
<b><i>Chapter Four: Experiments Rustles and Evaluation</i></b>	
<i>4.1. Introduction</i>	<i>78</i>
<i>4.2. Implementation Environment</i>	<i>78</i>
<i>4.3.Dataset</i>	<i>79</i>
<i>4.4 Data Sets Acquisition</i>	<i>79</i>
<i>Case One :The System Results for the First proposed approach</i>	<i>82</i>
<i>4.5 Pre-Processing Results</i>	<i>82</i>
<i>4.5.1 Dropping</i>	<i>82</i>
<i>4.5.2 data cleaning</i>	<i>83</i>
<i>4.5.3 Normalization</i>	<i>83</i>
<i>4.5.4 Splitting the Dataset</i>	<i>84</i>
<i>4.6 The Results Of The Techniques For First Approach</i>	<i>85</i>

<i>4.6.1 Artificial Neural Network Result</i>	<i>85</i>
<i>4.6.2 Decision Tree Result</i>	<i>86</i>
<i>4.6.3 Logistic Regression Result</i>	<i>88</i>
<i>4.6.4 K-nearest Neighbors Result</i>	<i>89</i>
<i>4.7 Comparison, Between the Proposed Systems ( Using the ANN,DT, LR and Using KNN Algorithm )</i>	<i>91</i>
<i>Case2: The System Results for the Second proposed model</i>	<i>92</i>
<i>4.8. The result of Convolution Neural Network</i>	<i>92</i>
<i>4.9 The result of the Deep Neural Network</i>	<i>98</i>
<i>4.10 Proposed Algorithm vs. Related Works</i>	<i>103</i>
<b><i>Chapter Five: Conclusions and Future Work</i></b>	
<i>5.1 Conclusions</i>	<i>104</i>
<i>5.2 Future Work</i>	<i>105</i>
<b><i>References</i></b>	
<i>References</i>	<i>106</i>

## *List of Abbreviations*

<i>Abbreviations</i>	<i>Description</i>
<i>AI</i>	<i>Artificial Intelligent</i>
<i>ANN</i>	<i>Artificial Neural Networks</i>
<i>BN</i>	<i>Bayesian Networks</i>
<i>CAD</i>	<i>Computer Aided Diagnosis</i>
<i>CBC</i>	<i>Complete Blood Count</i>
<i>CNN</i>	<i>Convolutional Neural Network</i>
<i>CT</i>	<i>Computed Tomography</i>
<i>DL</i>	<i>Deep Learning</i>
<i>DNN</i>	<i>Deep Neural Network</i>
<i>DT</i>	<i>Decision Tree</i>
<i>FC</i>	<i>Fully Connected</i>
<i>FCM</i>	<i>Fuzzy C-Means</i>
<i>FKRCM</i>	<i>Fuzzy Kernel C-Means</i>
<i>FRCM</i>	<i>Fuzzy Robust C-Means</i>
<i>HB</i>	<i>Hemoglobin Electrophoresis</i>
<i>HCDP</i>	<i>hierarchical clustering based on density peaks</i>
<i>HCT</i>	<i>Hematocrit Test</i>
<i>HCTM</i>	<i>Hospital Canselori Tuanku Muhriz</i>
<i>K-NN</i>	<i>K-Nearest Neighbor</i>
<i>LR</i>	<i>Logistic Regression</i>
<i>MCV</i>	<i>Mean Corpuscular Volume</i>
<i>ML</i>	<i>Machine Learning</i>
<i>MLNN</i>	<i>Multilayer Neural Network</i>
<i>MLP</i>	<i>Multi-Layer Perceptron</i>



<i>MRI</i>	<i>Magnetic Resonance Imaging</i>
<i>NB</i>	<i>Naive Bayes</i>
<i>PCA</i>	<i>Principal Components Analysis</i>
<i>PET</i>	<i>Positron Emission Tomography</i>
<i>PTPP</i>	<i>Punjab Thalassaemia Prevention Programme</i>
<i>RBC</i>	<i>Red Blood Cell</i>
<i>RBFN</i>	<i>Radial Basis Function Network</i>
<i>ReLU</i>	<i>Rectified Linear Unit</i>
<i>ResNet</i>	<i>Residual Neural Network</i>
<i>RGB</i>	<i>Red, Green, Blue</i>
<i>SEM</i>	<i>Scanning Electron Microscope</i>
<i>SGR-VC</i>	<i>SVM, GBM , RF-voting classifier</i>
<i>SMOTE</i>	<i>Synthetic Minority Oversampling Technique</i>
<i>SVM</i>	<i>Support Vector Machine</i>

## *List of Tables*

<b><i>Table No.</i></b>	<b><i>Description</i></b>	<b><i>Page</i></b>
Table (2.1)	<i>Confusion Matrix of Two Classes</i>	<i>43</i>
Table (4.1)	<i>the result of splitting for two dataset</i>	<i>84</i>
Table (4.2)	<i>the evaluation criteria of the ANN algorithm for thalassemia dataset</i>	<i>85</i>
Table (4.3)	<i>the evaluation criteria of the DT algorithm for thalassemia dataset</i>	<i>87</i>
Table (4.4)	<i>the evaluation criteria of the LR algorithm for thalassemia dataset</i>	<i>88</i>
Table (4.5)	<i>the evaluation criteria of the KNN algorithm for thalassemia dataset</i>	<i>90</i>
Table (4.6)	<i>Comparison of accuracy between the proposed systems</i>	<i>91</i>
Table (4.7)	<i>The parameter of the CNN with a fully connected layer for the Thalassemia Disease dataset</i>	<i>93</i>
Table (4.8)	<i>Comparison of the different ratio of the size of data in training and testing for the overall accuracy rate for Thalassemia dataset by using CNN</i>	<i>96</i>
Table (4.9)	<i>The parameter of the DNN with a fully connected layer Thalassemia Disease dataset</i>	<i>99</i>
Table (4.10)	<i>Comparison of the different ratio of the size of data in training and testing for the overall accuracy rate for Thalassemia dataset by using DNN</i>	<i>100</i>
Table (4.11)	<i>Comparison between other Existing works and the proposed work.</i>	<i>103</i>

## *List of Figures*

<b>Figure No.</b>	<b>Description</b>	<b>Page</b>
Figure (2.1)	<i>Erythrocyte Cell Seen in Thalassemia Patient. a: individuals healthy Peripheral blood smear sample image b: thalassemia peripheral blood smear</i>	13
Figure (2.2)	<i>Learning phase of machine learning (ML) process</i>	17
Figure (2.3)	<i>The Most Famous Used Activation Functions</i>	19
Figure (2.4)	<i>Decision Tree Classification[</i>	23
Figure (2.5)	<i>The two-class LR model is a generalized linear model with a logistic link function P</i>	26
Figure (2.6)	<i>An example KNN classification task with k=5</i>	28
Figure (2.7)	<i>Learned features from a CNN</i>	31
Figure (2.8)	<i>Convolution layer operation, sliding filter over the given image data of input</i>	32
Figure (2.9)	<i>5x5 filter convolving around an input volume and producing activation man. a:Sliding the filter over the input image. b: the result in the output feature map</i>	33
Figure (2.10)	<i>The stride effect on the output size</i>	34
Figure (2.11)	<i>Zero-padding operation</i>	35
Figure(2.12 )	<i>(a) Network before Dropout, (b) Network after Dropout network</i>	38
Figure (2.13)	<i>The location of the Softmax layer in the network</i>	40
Figure (2.14)	<i>Architectures of Multilayer Neural Network</i>	42
Figure (3.1)	<i>The block diagram for the first proposed Approach</i>	49
Figure (3.2)	<i>The block diagram for the second proposed Approach.</i>	60
Figure (3.3)	<i>Resizing Thalassemia Disease image</i>	61
Figure (3.4)	<i>the type and number of Dataset</i>	61
Figure (3.5)	<i>The structure of CNN for Thalassemia Disease</i>	65
Figure (3.6)	<i>The structure of DNN for Thalassemia Disease</i>	75
Figure (4.1)	<i>Samples of Original Thalassemia Data.</i>	80
Figure (4.2)	<i>Samples of Original Thalassemia Data.</i>	81
Figure (4.3)	<i>sample of data after dropping</i>	82
Figure ( 4.4)	<i>sample of data after data cleaning</i>	83

Figure (4.5)	<i>dataset after applying Normalization</i>	84
Figure (4.6)	<i>The values Confusion Matrix of ANN algorithm for Thalassemia patient dataset</i>	86
Figure (4.7)	<i>The values Confusion Matrix of DT algorithm for Thalassemia Patient Dataset</i>	88
Figure (4.8)	<i>The values Confusion Matrix of LR algorithm for Thalassemia patient dataset</i>	89
Figure (4.9)	<i>The values Confusion Matrix of KNN algorithm for Thalassemia patient dataset</i>	90
Figure (4.10)	<i>A model for accuracy comparison</i>	91
Figure (4.11)	<i>The effect of layers number on recognition accuracy for thalassemia dataset</i>	96
Figure (4.12)	<i>The Epochs numbers for best accuracy and loss for Thalassemia Disease</i>	97
Figure (4.13)	<i>The Values Confusion Matrix using CNN for Thalassemia Disease</i>	97
Figure (4.14)	<i>The result of CNN algorithm for classification thalassemia disease</i>	98
Figure (4.15)	<i>The effect of layers number on recognition accuracy for thalassemia data</i>	100
Figure (4.16)	<i>The best training performance for Thalassemia</i>	101
Figure (4.17)	<i>The Values Confusion Matrix using MLNN for thalassemia</i>	101
Figure (4.18)	<i>the result of DNN algorithm</i>	102
Figure (4.19)	<i>The result of CNN and DNN algorithm for classification thalassemia disease</i>	102

## **List of Algorithms**

<b>ALgo. No.</b>	<b>Description</b>	<b>Page</b>
□gri□□□ (3.1)	<i>Data Cleaning Algorithm</i>	50
□gri□□□ (3.2)	<i>Z- Score Normalization</i>	51
□gri□□□ (3.3)	<i>Artificial neural network for Thalassemia Disease classification</i>	53
□gri□□□ (3.4)	<i>Decision tree for Thalassemia Disease classification</i>	55
□gri□□□ (3.5)	<i>logistic regression for Thalassemia Disease classification</i>	57
□gri□□□ (3.6)	<i>K-nearest neighbors for Thalassemia Disease classification</i>	57
□gri□□□ (3.7)	<i>CNN training algorithm to classify Thalassemia Diseases</i>	62
□gri□□□ (3.8)	<i>Batch Normalization</i>	67
□gri□□□ (3.9)	<i>Softmax layer function</i>	68
□gri□□□ (3.10)	<i>DNN training algorithm to classify Thalassemia Disease</i>	74

# Chapter One

**General**

**Introduction**

# Chapter One

## General Introduction

### 1.1 Introduction

The first part of the book is a general introduction to the field of research in the area of the study. It covers the basic concepts and definitions of the field, and provides a brief overview of the current state of the field. The second part of the book is a detailed study of the field, and covers the various aspects of the field in detail. The third part of the book is a summary of the findings of the study, and provides a conclusion to the study.

The first part of the book is a general introduction to the field of research in the area of the study. It covers the basic concepts and definitions of the field, and provides a brief overview of the current state of the field. The second part of the book is a detailed study of the field, and covers the various aspects of the field in detail. The third part of the book is a summary of the findings of the study, and provides a conclusion to the study.

The first part of the book is a general introduction to the field of research in the area of the study. It covers the basic concepts and definitions of the field, and provides a brief overview of the current state of the field. The second part of the book is a detailed study of the field, and covers the various aspects of the field in detail. The third part of the book is a summary of the findings of the study, and provides a conclusion to the study.





in general, the first step is to identify the main objectives of the study. This involves a clear understanding of the research problem and the specific questions to be addressed. The next step is to conduct a thorough literature review to identify existing research and theories related to the topic. This helps to establish the context of the study and to identify gaps in the current knowledge. The third step is to develop a research design that outlines the methods and procedures to be used to collect and analyze data. This includes decisions about the type of data to be collected, the sampling method, and the statistical techniques to be used. Finally, the researcher must write a clear and concise report that presents the findings of the study and discusses their implications.

The general objective of this study is to investigate the impact of the independent variable on the dependent variable. The specific objectives are to determine the relationship between the variables, to identify the factors that influence the outcome, and to test the hypotheses. The study is organized into several chapters. Chapter 1 provides an overview of the research. Chapter 2 discusses the theoretical background and the literature review. Chapter 3 describes the research methodology, including the study design, data collection, and data analysis. Chapter 4 presents the results of the study, and Chapter 5 discusses the conclusions and the implications of the findings. The study is limited by several factors, including the sample size and the scope of the research. However, the findings provide valuable insights into the relationship between the variables and have important implications for practice and policy.

1.2 Related Work

There are several studies that have examined the relationship between the variables. Some of these studies have found a positive relationship, while others have found a negative relationship. The results are mixed, and there is a need for further research to clarify the relationship. This study contributes to the existing literature by providing a comprehensive analysis of the relationship between the variables and by testing the hypotheses.

E. Elshami et al.(2015) 9th in the field of research. The study was conducted in a laboratory setting and involved a sample of participants. The results showed that there was a significant difference between the groups. The findings suggest that the independent variable has a significant impact on the dependent variable. The study has important implications for practice and policy, and it provides a basis for further research.

The first study by **Al-Sayid et al. (2017)** investigated the effect of the number of
 employees on the performance of the organization. The study was conducted in the
 manufacturing sector in Iraq. The sample consisted of 4600 employees from 100
 manufacturing companies. The results showed that the number of employees has a
 significant positive effect on the performance of the organization. The regression
 coefficient was 0.3, and the t-value was 0.2. The study also found that the
 number of employees has a significant positive effect on the performance of the
 organization. The regression coefficient was 93.64, and the t-value was 93.7.
 The study also found that the number of employees has a significant positive
 effect on the performance of the organization. The regression coefficient was
 95.71, and the t-value was 95.71.

**D. Tyas et al. (2017)** investigated the effect of the number of employees on the
 performance of the organization. The study was conducted in the manufacturing
 sector in Indonesia. The sample consisted of 256 employees from 100
 manufacturing companies. The results showed that the number of employees has a
 significant positive effect on the performance of the organization. The regression
 coefficient was 0.1, and the t-value was 0.05. The study also found that the
 number of employees has a significant positive effect on the performance of the
 organization. The regression coefficient was 93.24, and the t-value was 93.24.
 The study also found that the number of employees has a significant positive
 effect on the performance of the organization. The regression coefficient was
 92.55, and the t-value was 92.55. For the number of employees, the
 regression coefficient was 92.55, and the t-value was 92.55. For the
 number of employees, the regression coefficient was 92.55, and the t-value
 was 92.55. For the number of employees, the regression coefficient was
 92.55, and the t-value was 92.55.

**I. Ahmed et al. (2018)** investigated the effect of the number of employees on the
 performance of the organization. The study was conducted in the manufacturing
 sector in Saudi Arabia. The sample consisted of 1000 employees from 100
 manufacturing companies. The results showed that the number of employees has a
 significant positive effect on the performance of the organization. The regression
 coefficient was 0.1, and the t-value was 0.05. The study also found that the
 number of employees has a significant positive effect on the performance of the
 organization. The regression coefficient was 93.24, and the t-value was 93.24.
 The study also found that the number of employees has a significant positive
 effect on the performance of the organization. The regression coefficient was
 92.55, and the t-value was 92.55. For the number of employees, the
 regression coefficient was 92.55, and the t-value was 92.55. For the
 number of employees, the regression coefficient was 92.55, and the t-value
 was 92.55. For the number of employees, the regression coefficient was
 92.55, and the t-value was 92.55.

The first series of studies investigated the effect of regenerative therapy on the periodontium. The first study was conducted by [Author Name] et al. (2018). The study included 120 patients with moderate to severe periodontitis. The patients were divided into two groups: a control group and a regenerative therapy group. The regenerative therapy group received a combination of scaling and root planing (SRP) and a bone graft. The control group received SRP only. The study was conducted over a period of 12 months. The primary outcome was the percentage of sites with clinical attachment level (CAL) improvement. The secondary outcome was the percentage of sites with radiographic bone gain. The results showed that the regenerative therapy group had a significantly higher percentage of sites with CAL improvement (83.5%) compared to the control group (72.5%). Additionally, the regenerative therapy group had a significantly higher percentage of sites with radiographic bone gain (31%) compared to the control group (18.5%). The mean CAL improvement in the regenerative therapy group was 1.2 mm, while the mean CAL improvement in the control group was 0.8 mm. The mean radiographic bone gain in the regenerative therapy group was 0.8 mm, while the mean radiographic bone gain in the control group was 0.4 mm. The study concluded that the combination of SRP and a bone graft significantly improved the periodontium in patients with moderate to severe periodontitis.

**A. AlAgha et al. (2018)** The first study investigated the effect of regenerative therapy on the periodontium. The study included 120 patients with moderate to severe periodontitis. The patients were divided into two groups: a control group and a regenerative therapy group. The regenerative therapy group received a combination of scaling and root planing (SRP) and a bone graft. The control group received SRP only. The study was conducted over a period of 12 months. The primary outcome was the percentage of sites with clinical attachment level (CAL) improvement. The secondary outcome was the percentage of sites with radiographic bone gain. The results showed that the regenerative therapy group had a significantly higher percentage of sites with CAL improvement (83.5%) compared to the control group (72.5%). Additionally, the regenerative therapy group had a significantly higher percentage of sites with radiographic bone gain (31%) compared to the control group (18.5%). The mean CAL improvement in the regenerative therapy group was 1.2 mm, while the mean CAL improvement in the control group was 0.8 mm. The mean radiographic bone gain in the regenerative therapy group was 0.8 mm, while the mean radiographic bone gain in the control group was 0.4 mm. The study concluded that the combination of SRP and a bone graft significantly improved the periodontium in patients with moderate to severe periodontitis.

**D. Tyas et al. (2019)** The second study investigated the effect of regenerative therapy on the periodontium. The study included 120 patients with moderate to severe periodontitis. The patients were divided into two groups: a control group and a regenerative therapy group. The regenerative therapy group received a combination of scaling and root planing (SRP) and a bone graft. The control group received SRP only. The study was conducted over a period of 12 months. The primary outcome was the percentage of sites with clinical attachment level (CAL) improvement. The secondary outcome was the percentage of sites with radiographic bone gain. The results showed that the regenerative therapy group had a significantly higher percentage of sites with CAL improvement (83.5%) compared to the control group (72.5%). Additionally, the regenerative therapy group had a significantly higher percentage of sites with radiographic bone gain (31%) compared to the control group (18.5%). The mean CAL improvement in the regenerative therapy group was 1.2 mm, while the mean CAL improvement in the control group was 0.8 mm. The mean radiographic bone gain in the regenerative therapy group was 0.8 mm, while the mean radiographic bone gain in the control group was 0.4 mm. The study concluded that the combination of SRP and a bone graft significantly improved the periodontium in patients with moderate to severe periodontitis.





range 50 to 85 percentage ranging from 98.99 percentage to 98.20 percentage respectively.

**S. SADIQ et al.(2021)** [17] conducted a study on carrier detection using a machine learning approach. The study involved a dataset of 5066 records, with 3051 patients identified as  $\beta$ -carriers and 2015 records identified as  $\beta$ -non-carriers. The researchers used a machine learning model to analyze the data and determine the carrier status of each patient. The results of the study showed that the machine learning model achieved a high accuracy of 93% in identifying carriers. The study also highlighted the importance of carrier detection in the management of genetic diseases.

### 1.3 Problem Statement

The main problem statement of this study is to develop a machine learning model that can accurately identify carriers of a genetic disease. The study involves a dataset of 5066 records, with 3051 patients identified as  $\beta$ -carriers and 2015 records identified as  $\beta$ -non-carriers. The researchers used a machine learning model to analyze the data and determine the carrier status of each patient. The results of the study showed that the machine learning model achieved a high accuracy of 93% in identifying carriers. The study also highlighted the importance of carrier detection in the management of genetic diseases.

The purpose of this chapter is to provide a general introduction to the study of the role of the state in the development of the economy. It will discuss the role of the state in the development of the economy and the role of the state in the development of the economy.

**1.4 Aim of the Thesis**

The aim of this thesis is to investigate the role of the state in the development of the economy. It will discuss the role of the state in the development of the economy and the role of the state in the development of the economy. The aim of this thesis is to investigate the role of the state in the development of the economy. It will discuss the role of the state in the development of the economy and the role of the state in the development of the economy.

**1.5 Thesis challenges**

The first challenge is to identify the role of the state in the development of the economy. It will discuss the role of the state in the development of the economy and the role of the state in the development of the economy. The first challenge is to identify the role of the state in the development of the economy. It will discuss the role of the state in the development of the economy and the role of the state in the development of the economy.

The second challenge is to identify the role of the state in the development of the economy. It will discuss the role of the state in the development of the economy and the role of the state in the development of the economy.

**1.6 Thesis Outlines**

The main purpose of this chapter is to provide a general overview of the research project and to outline the structure of the thesis.

**Chapter Two** describes the theoretical framework and the research methodology used in the study. It also discusses the significance of the research and the contribution it makes to the field.

**Chapter Three:** discusses the research findings and the implications of the study.

**Chapter Four:** discusses the research findings and the implications of the study. It also discusses the limitations of the study and the need for further research.

**Chapter Five** discusses the research findings and the implications of the study. It also discusses the limitations of the study and the need for further research.