A CONVOLUTIONAL NEURAL NETWORK BASED DETECTION SYSTEM FOR ACUTE APPENDICITIS DISEASE

TIMOTEWOS, MECHAL

http://hdl.handle.net/123456789/11704
Downloaded from DSpace Repository, DSpace Institution's institutional repository
BAHIR DAR UNIVERSITY
BAHIR DAR INSTITUTE OF TECHNOLOGY (BiT)
SCHOOL OF RESEARCH AND POSTGRADUATE STUDIES
FACULTY OF ELECTRICAL AND COMPUTER ENGINEERING

A CONVOLUTIONAL NEURAL NETWORK BASED DETECTION SYSTEM FOR ACUTE APPENDICITIS DISEASE

BY
MECHAL TIMOTEWOS

PROGRAM: COMPUTER ENGINEERING
MAIN ADVISOR: TESFA TESEGNE (PhD)
CO-ADVISOR: FIKRESELAM GARED (PhD)

Bahir Dar, Ethiopia
July, 2020
A Thesis

Submitted to The School of Research and Post Graduate Studies of

Bahir Dar Institute of Technology

In Partial Fulfilment of the Requirements for The Degree of

Master of Science in Computer Engineering

In The Faculty of Electrical and Computer Engineering.

By

Mechal Timotewos

Bahir Dar, Ethiopia

July, 2020
Bahir Dar University

Bahir Dar Institute of Technology (BiT)

Faculty of Electrical and Computer Engineering

A Convolutional Neural Network Based Detection System For Acute Appendicitis Disease

By

Mechal Timotewos Gibiso

Main Advisor: Tesfa Tegegne(Ph.D)

Co-Advisor: Fikreselam Gared (Ph.D)
DECLARATION

I, the undersigned student, declare that this thesis work which is entitled “A Convolutional Neural Network Based Detection System For Acute Appendicitis Disease” is my own work and any form of documents used in this work have been acknowledged by reference.

Name of the student Mechaltimotewos Signature

Date of submission: 16/11/2012

Place: Bahir Dar

This thesis has been submitted for examination with my approval as a university advisor.

Advisor’s Name: Tesfa Telegne (PLD)

Advisor’s Signature:

© 2020 Mechaltimotewos
BAHIR DAR UNIVERSITY
BAHIR DAR INSTITUTE OF TECHNOLOGY (BiT)
SCHOOL OF RESEARCH AND POST GRADUATE STUDIES
FACULTY OF ELECTRICAL AND COMPUTER ENGINEERING

THESIS APPROVAL SHEET

The following graduate faculty members certify that this student has successfully presented the necessary written final thesis and oral presentation for partial fulfillment of the thesis requirements for the Degree of Master of Science in Computer Engineering.

Approved by:

Advisor:
Tesfa Tesfahun (PhD)  
Name  Signature  Date  16/11/2012 E.C

External Examiner:
Abebe Tesfahun (PhD)  
Name  Signature  Date  15/11/2012 E.C

Internal Examiner:
Abraham Debark  
Name  Signature  Date  16/11/2012 E.C

Chair Holder:
Abebe Tesfahun  
Name  Signature  Date  16/11/2012 E.C

Faculty Dean:
Tewodros Sera Workineh  
Name  Signature  Date  16/11/2012 E.C
ACKNOWLEDGEMENT

First and foremost, I would like to thank God for all his support and care throughout my life. I offer my special thanks to my advisor Dr. Tesfa Tegegne and my co-adviser Dr. Fikreselam Gared for their support and guidance in all steps of the thesis work.

I would also like to offer my sincere thanks to my beloved wife Merry and my daughter Abigya for their unlimited patience and understanding.

I am most grateful to all the people who were in my side during the thesis work and their comments are appreciated.
ABSTRACT

Medical industry is one of the well-known sectors where predictions are very essential. This research study encourages the use of artificial intelligence for decision-making processes in the medical field for detecting acute appendicitis disease. Acute appendicitis is one of the most common abdominal pain that leads to emergency surgery. Currently, ultrasound has gained high acceptance for diagnosis of patients with severe abdominal pain because it can have no any radiation, cost effective, and is easily available.

The main objective of this study is to develop a Convolutional Neural Network Based Detection System for Acute Appendicitis Disease.

The existing acute appendicitis diagnosis systems mainly focus on the clinical and laboratory findings. But now a day, these results lacks acceptance due to they result in false diagnosis. To minimize this, results from imaging modalities have gained high acceptance worldwide. To help the physicians in decision making process for diagnosis acute appendicitis from imaging modalities specifically the ultrasound images, an intelligent model that can diagnosis acute appendicitis from ultrasound image is very important.

The image datasets which have been used for this study were collected from internationally available sources (pediatrics, 2011) (appendicitis during pregnancy, 2013) (encyclopaedia, n.d.) (abdomen and retroperitoneum, 2012) (Infant with perforated appendicitis, 2014). The Gaussian filter and histogram equalization techniques were applied in order to remove the noises in the image and increase the brightness of the images respectively. The proposed model is trained and validated for both pre-processed and unprocessed data. The model is tested with test data and the results obtained are used to evaluate the performance of the model.

The performance of the proposed model is evaluated based on accuracy, recall and specificity and the results obtained from the experimental analysis were presented. The proposed model achieved an accuracy of 99.16 for pre-processed data and 95.18 for unprocessed data. The performance of the proposed model is compared with pre-trained models and the result shows that our proposed model is a good choice than pre-trained ones.

Keywords: Appendicitis, Convolutional Neural Network, Ultrasound
# TABLE OF CONTENTS

DECLARATION .......................................................................................................................... i

ACKNOWLEDGEMENT .......................................................................................................... ii

ABSTRACT ................................................................................................................................ iv

LIST OF ABBREVIATIONS ....................................................................................................... ix

Chapter 1 .................................................................................................................................. 1

Introduction ............................................................................................................................... 1

1.1. Background ....................................................................................................................... 1

1.2. Motivation .......................................................................................................................... 4

1.3. Statement of Problem ........................................................................................................ 5

1.4. Objectives .......................................................................................................................... 6

1.4.1. General Objective ......................................................................................................... 6

1.4.2. Specific Objectives ....................................................................................................... 6

1.5. Scope of the study .............................................................................................................. 7

1.6. Methodology ..................................................................................................................... 7

1.7. Significance of The Study ................................................................................................. 8

1.8. Thesis Organization .......................................................................................................... 8

Chapter 2 .................................................................................................................................. 9

Literature Review .................................................................................................................... 9

2.1. Overview ........................................................................................................................... 9

2.2. General Approaches for Diagnosing Appendicitis Disease ............................................. 9
2.2.1. Diagnosis Based on Clinical and Laboratory Findings ..................................10
2.2.2. Diagnosis Based on Imaging Tests ..........................................................10
2.3. Digital Image processing ..............................................................................12
2.4. Neural Networks .........................................................................................12
  2.4.1. Working principles of CNN Algorithm .....................................................13
  2.4.2. CNN Architectures ..................................................................................17
2.5. Related Works .............................................................................................20
  2.5.1. Related Works Based on Clinical and Laboratory Results ......................20
  2.5.2. Related Works Based On Imaging Modalities .........................................22
2.6. Chapter Summary ........................................................................................24

Chapter 3 ...........................................................................................................25

METHODOLOGY .................................................................................................25
  3.1. Overview .....................................................................................................25
  3.2. System Architecture ...................................................................................25
  3.3. Proposed Model Architecture ....................................................................27
  3.4. Dataset Preparation ....................................................................................29
    3.4.1. Data Augmentation ...............................................................................29
  3.5. Loading and Pre-processing Images ...........................................................30
  3.6. Feature Extraction ......................................................................................33
  3.7. Training and Classification .........................................................................33
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.8.</td>
<td>Performance Evaluation Metrics</td>
<td>34</td>
</tr>
<tr>
<td>3.8.1.</td>
<td>Confusion Matrix</td>
<td>34</td>
</tr>
<tr>
<td>3.8.2.</td>
<td>Classification Accuracy</td>
<td>35</td>
</tr>
<tr>
<td>3.8.3.</td>
<td>Sensitivity/Recall</td>
<td>35</td>
</tr>
<tr>
<td>3.8.4.</td>
<td>Specificity</td>
<td>36</td>
</tr>
<tr>
<td>3.8.5.</td>
<td>Precision</td>
<td>36</td>
</tr>
<tr>
<td>3.8.6.</td>
<td>F-measure</td>
<td>37</td>
</tr>
<tr>
<td>3.9.</td>
<td>Developing tools</td>
<td>37</td>
</tr>
<tr>
<td>3.10.</td>
<td>Training The Model</td>
<td>38</td>
</tr>
<tr>
<td>3.11.</td>
<td>Model parameters</td>
<td>38</td>
</tr>
<tr>
<td>3.12.</td>
<td>Chapter Summary</td>
<td>39</td>
</tr>
<tr>
<td>4.1.</td>
<td>Overview</td>
<td>40</td>
</tr>
<tr>
<td>4.2.</td>
<td>Output of Image Pre-processing</td>
<td>40</td>
</tr>
<tr>
<td>4.3.</td>
<td>Summary of proposed model architecture</td>
<td>41</td>
</tr>
<tr>
<td>4.4.</td>
<td>Model performance for different learning rates</td>
<td>42</td>
</tr>
<tr>
<td>4.5.</td>
<td>Model performance for different train test split ratio</td>
<td>42</td>
</tr>
<tr>
<td>4.6.</td>
<td>Implementation</td>
<td>43</td>
</tr>
<tr>
<td>4.6.1.</td>
<td>Image Classification on Pre-Processed Data</td>
<td>43</td>
</tr>
</tbody>
</table>
LIST OF ABBREVIATIONS

AA  Acute appendicitis
AI  Artificial Intelligence
CNN Convolutional Neural Network
CT  Computed Tomography
ED  Emergency Department
FN  False Negative
FP  False Positive
GB  Giga Byte
GPU Graphical processing unit
HDD Hard Disk Drive
HE  Histogram Equalization
IDE Integrated Development Environment
ML  Machine Learning
MRI Magnetic Resonance Image
NN  Neural Network
RAM Random Access Memory
ReLU Rectified Linear Unit
RGB Red, Green, and Blue
TB  Tera Byte
TN  True Negative
TP  True Positive
US  Ultrasound
2D  Two dimensional
LIST OF FIGURES

Fig 2. 1: Over all architecture of a neural network .................................................13
Fig 2. 2: Convolution operation in an input image ..................................................14
Fig 2. 3: Average and Max pooling operations ......................................................15
Fig 2. 4: Flattening operation .............................................................................17
Fig 2. 5: Fully connected layers .......................................................................17
Fig 2. 6: Basic architecture of LeNet-5 ...............................................................18
Fig 2. 7: Basic architecture of AlexNet ...............................................................19
Fig 2. 8: Basic architecture of VGGNet ..............................................................19
Fig 3. 1: Detailed system architecture ...............................................................26
Fig 3. 2: Proposed model architecture ..............................................................28
Fig 3. 3: (a) Input image of AA +ve patient (b) Input image of AA –ve patient .......31
Fig 3. 4: Confusion matrix for binary class ......................................................34
Fig 4. 1: Output image .....................................................................................40
Fig 4. 2: Proposed model summary ..................................................................41
Fig 4. 3: Model Accuracy on processed data ..................................................43
Fig 4. 4: Model Loss on processed data ............................................................44
Fig 4. 5: Model Accuracy on un-processed data .............................................45
Fig 4. 6: Model Loss on un-processed data ......................................................45
Fig 4. 7: Model Accuracy on VGG19 model for processed .........................46
Fig 4. 8: Model Loss on VGG19 model for processed .................................47
Fig 4. 9: Model Accuracy on InceptionNetV3 model for processed .............48
Fig 4. 10: Model Loss on InceptionNetV3 model for processed ..................48
LIST OF TABLES

Table 3. 1: Experimental specification for the experiment........................................37
Table 4. 1: Model performance on learning rates ..................................................42
Table 4. 2: Model performance for different data split ..........................................42
Table 4. 3: Model hyper parameters........................................................................43
Table 4. 4: Confusion matrix on pre-processed data..............................................44
Table 4. 5: Performance summary pre-processed data.........................................44
Table 4. 6: Confusion matrix on unprocessed data...............................................46
Table 4. 7: Performance summary on un-processed data.....................................46
Table 4. 8: Model test loss and test accuracy.......................................................47
Table 4. 9: Performance summary on InceptionNet..............................................49
**LIST OF SYMBOLS**

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Input image</td>
</tr>
<tr>
<td>K</td>
<td>Convolution kernel</td>
</tr>
<tr>
<td>Z</td>
<td>Pooled feature map</td>
</tr>
<tr>
<td>( F^k )</td>
<td>Input feature map</td>
</tr>
<tr>
<td>( T^k )</td>
<td>Transformed output</td>
</tr>
<tr>
<td>( \Sigma )</td>
<td>Summation</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

1.1. Background

Artificial intelligence, which is also called a machine intelligence is one of an emerging technology that contains mechanisms that behave in a way considered intelligent if a human would act so. It has been used successfully to solve different real-world problems in almost all domains, that are connected to a large variety of applications, that need to have an intelligent behavior. These intelligent behaviors can be related to the capacity of perception and learning, to the ability to make decisions for different applications like gaming, self-driving cars, medical diagnosis or to other aspects that involve intelligent reasoning (McCarthy, 2007).

In general, a decisional system is mainly not intended to replace the physicians in medical field (however, it would be impossible). It only makes suggestions which supports physicians for further actions and interpretations. In medical sector, the final medical decision belongs to humans. But some autonomous intelligent systems are useful for two important reasons (McCarthy, 2007):

- to avoid simple and usual tasks that are time consuming and overload the medical stuff with a lot of activities that could be easily performed by a machine (for example, to verify that whether certain patients are able or not to join certain company for job, or to predict the effectiveness of some drugs for certain medical diagnosis from previous patients’ history);
- to point out the medical conditions of patients which can be difficult to be detected by humans; a suggestion or an alert can make the difference between life and death (for example, to detect and diagnosis some cancerous cells from medical images such as x-ray, US, MRI or CT scan, or to predict the risk of getting infected by some genetic disease from the clinical information and other related background history of the patients).
Acute appendicitis disease is one of the most common conditions which can be treated by emergency operation, which is removal of the appendix. It is an inflammation of the appendix and there is no universally accepted effective medical therapy that patients can consult the doctor and get treatment promptly. Most patients of acute appendicitis recover with less or no difficulties as long as they get appropriate treatment early. If the treatment for AA is delayed, the case might become complex causing the appendix to burst, causing severe infection which eventually leads to death (Devi, Design and Development of Automatic Appendicitis Detection System using Sonographic Image Mining, 2012).

It is considered as one of the most common cause of acute abdominal pain that requires immediate surgical intervention. It has been reported and believed by different scholars that the lifetime risk for appendicitis for males is 8.6% and for females 6.7% (R. Jeffry, Acute Appendicitis: High-Resolution Real-Time US Findings, 1987).

A research study conducted in Zewuditu Memorial Hospital, Addis Ababa, Ethiopia, shows that Acute appendicitis was the commonest cause of acute abdominal emergency in children (Awetash, 2014). It was more prevalent in children aged 10-14 years. It is also well known from many clinical research papers that traditionally, appendicitis disease can be diagnosed solely based on clinical symptoms and signs, and later as time goes, the diagnosis included results of inflammatory laboratory variables such as Leukocytes Count, Urinalysis, and neutrophils. This practice of diagnosing acute appendicitis disease led to a false positive diagnosis (negative appendectomy) rates which varies in the range of 15-30% (R. Jeffry, 1987).

The development of imaging modalities, such as ultrasound, magnetic resonance image, and especially that of computed tomography, has enabled more accurate diagnostics with a significant decrease in false positive diagnoses, which has led to lower rates of negative appendectomies (Devi, 2012). It is a condition which can be characterized by an inflammation of the vermiform appendix, which is considered as one of the most common causes of the acute abdomen. The AA is one of the most frequent indications for an emergent abdominal surgical procedure all over the world (F. Charles, 2015).

The incidence of appendicitis is highest between the ages of 10 and 19 years, and men are more likely to develop appendicitis than women (Sammalkorpi).
Among the imaging based diagnostic tools available worldwide, Ultrasound (US) based diagnosis, as one of the most used imaging diagnoses, has been recognized as a powerful and ubiquitous and widely accepted screening and diagnostic tool for AA diagnosis. In particular, US imaging is widely used in children’s and pregnant women’s in most of the diagnosis procedures worldwide due to its many advantages like its safety to patients, low cost, non-invasive nature, real-time display, operator comfort and simplicity, and less operator experience when compared to other imaging modalities. Over a long period of time, it has been demonstrated by different researchers that US based diagnosis has several major advantages over other medical imaging modalities such as X-ray, magnetic resonance imaging (MRI), and computed tomography (CT). This is because of the benefits of its non-ionizing radiation, simplicity and portability of the device, ease of accessibility, and cost effectiveness for patients. However, US also presents unique challenges over other imaging modalities, such as low imaging quality caused by noise and artifacts, high dependence on abundant operator or diagnostician experience, and high inter- and intra-observer variability across different institutes and manufacturers’ US systems (Liu, 2019).

Abdominal pain and vomiting were mainly considered as the most frequent symptoms (100% and 79% respectively) whereas tenderness and guarding were the most frequent clinical signs found (83% and 61% respectively) for diagnosis of acute appendicitis (Birhanu, 2007).

From the imaging modalities available worldwide now a day, Ultrasound imaging modality is a widely available and inexpensive modality for any abdominal pain diagnosis due to its potential for highly accurate imaging in the patient suspected to have acute appendicitis. Although good skill and experience of an expert is an important factor in all ultrasound examinations, it has particular importance in the examination of the patient with right-lower-quadrant pain. Nonetheless, the criteria for the Ultrasound-based diagnosis of acute appendicitis are well established and reliable (E. Gaensler, 1999).

To date, US imaging for suspected Acute Appendicitis is performed world-wide by radiologists and many other professionals with other medical subspecialties, with or without the support of sonographers. However, in all imaging-based disease diagnosis procedures, the accurate detection and prediction of a disease outcome from MRI, CT and
US image is a challenging task that relies on the availability of expert radiologists. As a result, different Machine learning algorithms are becoming a popular tool for decision support systems for physicians in medical fields. These machine learning algorithms have the capability to discover and identify different patterns and relationships between the images from complex datasets, while they are able to effectively predict future outcomes of a particular problem.

The below listed symptoms indicates ultrasound signs and symptoms that indicates the presence of acute appendicitis disease (G. Mostbeck, 2016).

- When touched by the physician, the appendix part becomes non-compressible
- The appendix diameter becomes more than 6mm
- The thickness of the appendix becomes >=3mm
- Appendicolith: hyperechoic with posterior shadowing
- Local abscess formation occurs around the appendix

Diagnostic scoring which has been used for so long in the diagnosis of acute appendicitis was originally invented before the era of modern imaging technologies as an independent diagnostic tool. It has therefore often been simply investigated in the surgical literature as an alternative diagnosis means to imaging (A. Peter, 2015). However, both the diagnostic scoring method and imaging modality should optimally be used as complementary methods in a diagnostic algorithm. The aim is to achieve accurate diagnosis in the healthcare system with minimal risks, delays, and costs in a standardized manner independent of the experience level of the clinician and other experts. Lately, diagnostic scoring has been included in consensus guidelines of diagnosis of appendicitis (R. Ramon, 2016).

1.2. Motivation

Now days the field of artificial intelligence is becoming an interesting field and humans are using AI in every aspect of life including military, industry zones, medical diagnosis and treatment, etc. The motive for this research is to support physicians in decision making process during diagnosis of acute appendicitis by minimizing the labour and time required. To automate the medical field, we are motivated to develop an intelligent system that
contributes a lot for the medical automation system by providing the disease diagnosis system specifically the acute appendicitis disease diagnosis from US images.

Therefore, it is very important to implement automated diagnosis techniques in medical field that supports physicians in decision making process and also to avoid the challenges that results due to lack of an expert in the field.

1.3. Statement of Problem

Acute appendicitis is one of the most difficult diseases to diagnose clinically, even for experienced surgeons. Several methods, such as diagnosis based on clinical scoring system and laboratory tests, have been used to increase the accuracy of appendicitis diagnosis. However, these methods have not shown stable performances, including accuracy, sensitivity, and specificity, due to complex clinical protocols (Kim, 2015).

(R. Balu, 2011) Decision support systems for physicians are very crucial. Existing systems which support physicians in decision making process specifically in diagnosis of acute appendicitis has achieved good performance. However, they have some drawbacks which are discussed in literature review. The existing studies are mainly conducted on diagnosing acute appendicitis from clinical and laboratory data only for their designed system and currently, any diagnosis of the acute appendicitis can be done with the use of imaging modalities such as US, MRI or CT (G. Mostbeck, 2016). In addition to that, the performance of the existing systems is not satisfactory and needs to be improved.

Although there is one systems which is designed to diagnosis acute appendicitis using medical images specifically the CT scan images, it still encountered many drawbacks such as it only extracting part of the appendicitis while not considering other parameters surrounding the appendicitis like thickness of the appendicitis, even the diameter variation between different ages, perforation around the appendicitis, Local abscess formation and other fluid around the appendicitis which should be considered too. In addition to that CT images for the diagnosis can be used in rare cases due to its cost, and mainly its side effects like radiation (A. Walid, 2019) (R. Balu, 2011) (K. Baek, 2015) (M. Tez, 2008).

This study proposes a new approach for the diagnosis of acute appendicitis from ultrasound images by considering the parameters which are not considered in previous works. In
addition to that, there is no any system which is developed to diagnosis acute appendicitis from US image.

The novelty of our research work is that it solves the gaps encountered in previous studies like studying the surrounding parameters of appendicitis, developing a new model which is suitable for patients in all ages and circumstances (like pregnancies) in which other models fail to work.

This study attempts to address the following questions:

- To what extent do deep learning algorithms can diagnosis acute appendicitis from US image?
- By what extent do image pre-processing can increase the performance of the system?

1.4. Objectives

1.4.1. General Objective

The general objective of this research study is to develop A Convolutional Neural Network Based Detection System for Acute Appendicitis Disease.

1.4.2. Specific Objectives

To achieve the goal mentioned in the general objective, the following issues are addressed:

- Existing acute appendicitis diagnosis models/methods were examined.
- Factors that can affect the diagnosis of acute appendicitis disease were analysed.
- Acute appendicitis suspected patient history and related US images from different repositories were collected
- A prototype to diagnosis acute appendicitis is developed.
- The proposed acute appendicitis model is evaluated
- The performance of pre-trained models (VGG and InceptionNet) is evaluated
- The accuracy and loss of the developed acute appendicitis classification model is compared with other pre-trained models.
1.5. Scope of the study

To achieve the mentioned objectives above, this study is in topic areas of image resizing to make all images of fixed size, contrast enhancement to increase the contrast of the image for further processing and noise removal to remove undesired distortions from the image. The main task of this research work was developing a convolutional neural network based diagnosis model for AA from US images. In addition to that, the same dataset and hyper parameters was used to evaluate the performance of pre-trained models and the performance of the proposed model will be compared with other pre-trained models.

1.6. Methodology

This study employs an experimental research methodology. The methodology consists of the following steps:

- **Literature Review**

  The literature review part covers the relevant works related to this study. In this section, all related studies will be investigated in detail, and the contribution and drawbacks of each study will be presented in detail.

- **Data Collection**

  The dataset used in this study was collected from different internationally available repositories. The dataset that was used for this study are ultrasound images and the image pre-processing operations were applied to remove the noise and increase the contrast on the collected dataset.

- **Developing Tools**

  **Hardware tools:** A Desktop computer and virtual google colaboratory GPU was used.

  **Software tools:** The software tools that were used for this study are Mat Lab 2017a, anaconda IDE and python programming language.

- **Research Design** – The research design generally involves three phases. These are training phase where the proposed model is trained with the input image features and corresponding labels, Validation phase where the trained model is validated on
some amount of data and Testing phase where the trained model is tested with new unseen data to check its ability on performing the classification task.

- **Evaluation**

The performance of the new developed model was evaluated using internationally accepted performance metrics such as accuracy, precision, sensitivity, specificity and F Score measures.

1.7. **Significance of The Study**

It is obvious that manual analysis of any task is tedious, tiresome, vulnerable to error. To avoid the problems at all or minimize them to a high extent, it is very essential to have intelligent systems that behave in a way human do. So, having an intelligent model can provide the following benefits:

- Will help the medical world by easily detecting the acute appendicitis disease to support clinicians on decision making.
- Can solve the problems with the existing systems and previous studies by considering other parameters which are not yet considered.
- Can reduce negative appendectomy which results due to false interpretation (false positive, true negative) of the result.
- Helps to make better decisions
- Make the diagnosis of acute appendicitis disease diagnosis system simple and reliable.

1.8. **Thesis Organization**

The remaining part of this study is organized as follows. Chapter 2 reviews relevant literatures, Chapter 3 explains clearly about the methodology used for this research work. Chapter 4 shows Experimental results and analysis carried out throughout the developing process of the proposed model. Finally, Chapter 5 concludes the research work and points out future work directions.
CHAPTER 2

LITERATURE REVIEW

2.1. Overview

Recent technological advancements in almost all fields such as deep learning and the abundance of large data have enabled algorithms to surpass the performance of medical professionals in a wide variety of medical imaging tasks, including disease detection, classifications based on different category and risk factor identification for the disease (R. Pranav).

Researchers have worked and still working on different machine learning algorithms for disease diagnosis tasks like classification, risk factor prediction, identification and level of disease prediction. In (Kumar, 2018), used a machine learning techniques for prediction of kidney disease from clinical data and in (alii, 2016) used k-nearest neighbor (k-NN) and support vector machine (SVM) classifiers for predicting the presence of diabetes in patients. The researchers have been accepted and agreed on the same idea that different machine-learning algorithms work well in diagnosis of different diseases and this is because the algorithms have the ability to extract different patterns and features from raw data and predict future outcomes (Maruf, 2017).

There are existing approaches which can be used for diagnosing appendicitis disease which have been developed to increase the speed and accuracy of the diagnosis process, and to decrease the negative laparotomies (false positive results). So, this chapter briefly discusses the existing related works in detail with their contribution and limitation as well.

2.2. General Approaches for Diagnosing Appendicitis Disease

In diagnosing AA disease, almost all medical professionals suspect the diagnosis of appendicitis based on the symptoms, medical history, and a physical examination of suspected patients. But now a day, any health care professional can confirm the diagnosis of the disease with the help of imaging modalities such as an ultrasound, MRI examination,
or CT scan (Diagnosis of appendicitis). There are mainly two approaches which can be used to diagnosis acute appendicitis. These are:

2.2.1. Diagnosis Based on Clinical and Laboratory Findings

- **Medical History**

Almost all health care professional will diagnosis AA by asking the suspected patient’s specific questions like the period or length of abdominal pain, the exact location and severity of the pain, other medical conditions, previous illnesses, and other (Diagnosis of appendicitis).

- **Physical Exam**

In addition to the medical history of the suspected patients, the physicians may be interested to know details about the pain in patient’s abdomen to diagnose appendicitis correctly. They will assess the pain by touching or applying pressure to specific areas of the patient abdomen specifically the right lower abdomen part if he/she suspect that the disease is appendicitis (Diagnosis of appendicitis).

- **Laboratory Tests**

The medical history and physical examination of AA suspected patients is not enough for making a better decision. So medical professionals need some laboratory test result of the patients to help them confirm the diagnosis of appendicitis or find other causes of abdominal pain. This laboratory test includes blood test, urinalysis, and pregnancy test for women since vomiting is common for pregnancy and appendicitis (Diagnosis of appendicitis).

2.2.2. Diagnosis Based on Imaging Tests

Most of the time it is a common approach for health professionals to use imaging tests to confirm the diagnosis of appendicitis or find other causes of pain in the abdomen of suspected patients’ (Diagnosis of appendicitis):

- **Abdominal Ultrasound** – From the results of abdominal ultrasound images of suspected acute appendicitis patients, a health care professionals or a
radiologists reviews the images, which can show signs of (Diagnosis of appendicitis):

- A diameter of the appendix
- Local abscess around the appendix
- Inflammation of the appendix
- Perforation around the appendix
- others

A common approach for almost all health care professionals is to use an ultrasound test as first imaging modality on suspected patients for possible appendicitis in children, young adults and pregnant women.

- **Magnetic Resonance Imaging (MRI)**

In AA diagnosis process, MRI examination will not be the first choice due to its side-effects. MRI machines which can be used to produce detailed picture of patients’ uses a radio waves and magnets without using x-rays (Diagnosis of appendicitis). As like US image analysis, medical professionals’ assess the below signs if they suspect that the disease is AA (Diagnosis of appendicitis).

- A diameter of the appendix
- Local abscess around the appendix
- Inflammation of the appendix
- Perforation around the appendix
- Others cause of abdominal pain

Most medical professionals use an MRI examination as a good and alternative for diagnosis when compared to CT scan.

- **CT Scan** – CT scan tests mainly uses x-rays and computer technology to produce images from body organs of any patients. So, if the patient is suspected of acute appendicitis, the medical professionals’ assess the below listed signs from the CT scan image (Diagnosis of appendicitis).

- A diameter of the appendix
• Local abscess around the appendix
• Inflammation of the appendix
• Perforation around the appendix
• Others cause of abdominal pain

2.3. Digital Image processing

Digital image processing is a technique where some operations are performed on an input image in order to get an enhanced image or extract some other useful information from the image. It involves a sequence of steps listed below:

• **Image Acquistion** – In this step, the input image which is collected from different sources is scaled to the same size which is usually.
• **Image Pre-processing** – The aim of image pre-processing operation is to improve the image data by removing unwanted distortions from the images or enhance some image features which are important for further processing operations. In this step the noise removal and contrast enhancement techniques are applied which will be explained in chapter 3.

2.4. Neural Networks

A neural network represents a series of machine learning algorithms, which is built on the principle of the organization and functioning of human brain and function as a biological neuron. This networks basically consists of individual units called neurons and these neurons are located in groups to form layers. In neural networks, neurons in one layers are connected to one or more neurons in the other layers. In neural networks, the data which has to be processed comes from the input layers, passes through different hidden layers and goes to output layers for decision.

Each individual neuron in a neural networks performs a specific mathematical operation on the data it receives and transmits the modified data to all other neurons to which it is connected to (Image classification with CNN). The overall architecture of a neural network is presented below in Figure 2.1.
Among different existing deep learning algorithms, the convolutional neural network algorithm is used in our study and its working principle is described below.

### 2.4.1. Working principles of CNN Algorithm

In general, convolutional neural networks basically consist of an input layer, hidden layers and output layer. The input to our model in our case is a two dimensional image pixels, the hidden layers usually consist of convolutional layers, ReLU layers, pooling layers, and fully connected layers, and the output layer consists of possible outcomes of the CNN model (Common architectures in CNN).

Almost all CNN architectures follow the same general design principles. They all follow the principle of successively applying one or more convolutional layers to the input, and then periodically pooling or down sampling the spatial dimensions while increasing the number of feature maps (Common architectures in CNN).

There are different layers (components) that can be used in a typical convolutional neural network architecture. The basic CNN layers that are used in our study are discussed below:
• **Convolutional Layer**

The convolutional layer is the main component of a CNN algorithm and it is basically composed of a set of convolutional kernels which works by splitting the input image into small slices which is commonly called as receptive fields. This technique helps us to extract different important features of an image. During the convolution operation, the specified kernel convolves with the input images using a specific set of weights, by multiplying its elements with the corresponding elements of the image receptive field (Bouvrie, 2006). In general, Convolutional layers convolve the input and pass its result to the next layer. Generally, Convolution operation can be expressed mathematically as follows:

\[ f_1^k(p, q) = \sum_i i_c(x, y) \cdot e_i^k(u, v) \]  

(2.1)

where, \( i_c(x, y) \) is an element of the input image tensor \( I_c \), which is element wise multiplied by \( e_i^k(u, v) \) index of the \( k^{th} \) convolutional kernel \( K_i \) of the \( i^{th} \) layer.

![Fig 2. 2: Convolution operation in an input image](image)

Where I is the input image and K is the convolution kernel

• **Pooling Layer**

The basic working principle of a pooling layer is that it slides a two-dimensional filter over the input features and outputs a summary of the features. Generally, in pooling layers, the application of pooling operations reduces the dimensions of the data. Although there are different types of pooling layers that can be applied in CNNs, the most commonly used pooling layer for CNNs is Max Pooling and this is because it is vastly superior for capturing
invariances in images. Max pooling operation is all about taking the maximum value at each point in the image. This gets rid of 75% of the information that is not the feature (Common architectures in CNN).

In pooling operations, the output feature maps of the convolutional operation can occur in many different locations in the image. Once we extract the image features using pooling operation, the exact locations of the input image is not worthy that much as long as its approximate positions relative to others is preserved well. Pooling operation takes minimum, maximum, average or sum of similar information in the neighborhood of the receptive field and outputs the dominant response within this local region (C.-Y. Lee, 2016).

\[ Z^k_l = g_p(F^k_l) \] (2.2)

Where \( Z^k_l \), represents the pooled feature map of \( l^{th} \) layer for \( k^{th} \) input feature map \( F^k_l \), and \( g_p(.) \) defines the type of pooling operations.

Fig 2. 3: Average and Max pooling operations

- **Activation Function**

An activation function is a type of function that can be used to get the output of a node. It usually serves as a decision function and helps in learning of complex patterns in CNNs. An activation function of a convolved feature-map is mathematically defined as:

\[ T^k_l = g_a(F^k_l) \] (2.3)

15
Where $F_l^k$, is an output of a convolution which is assigned to activation function $g_a(.)$ that adds non-linearity and returns a transformed output $T_l^k$, for $l^{th}$ layer.

In different literatures, different types of activation functions like sigmoid, tanh, maxout, SWISH, ReLU, and variants of ReLU such as leaky ReLU, ELU, and PReLU are used to train non-linear combination of features. However, for CNN classification algorithms, ReLU and its variants are preferred as they help in overcoming the vanishing gradient problem (C. Nwankpa, 2018.) (Hochreiter, 1998).

- **Dropouts**

The Dropout is one of the widely used technique for controlling the capacity of a neural networks to prevent overfitting. It introduces regularization within the network which usually prevents overfitting with huge amount of training data. It is believed that dropouts ultimately improve generalization by randomly skipping some units or connections with a certain probability $p$ during training and uses all activations by scaling them by probability $p$ during testing.

In convolutional neural network architectures, different connection layers that learns a non-linear relation are sometimes co-adapted, which causes overfitting (G. E. Hinton, 2012). These process of randomly removing some units during training and testing phase can reduce the thinned network architecture, and finally one representative network is selected with small weights. This selected architecture is then considered as an approximation of the overall proposed networks (N. Srivastava, 2014).

- **Flattening Layers**

In flattening layer, the pooled feature maps are converted to a column vector and by doing so, we come up with a single long vector of input data that will be sent to the artificial neural network for further processing. The flattening layer operation is represented in the below figure.
• **Fully Connected Layers**

In neural networks, fully connected layers connect every neuron in one layer to every neuron in another layer (*Common architectures in CNN*). These layers are mostly used at the end of the network for the purpose of classification. Unlike pooling and convolution operations, it is a global operation. It takes input from feature extraction stages and globally analyses output of all the preceding layers (*M. Lin, 2013*).

2.4.2. **CNN Architectures**

These different CNN architectures that can be followed during the development of a CNN model can serve as an excellent feature extractor which can be used for different tasks like
image classification, object detection, image segmentation, and many other more advanced tasks. Some of the commonly used architectures for convolutional networks are (Common architectures in CNN):

- **LeNet-5**

This model was first developed to identify handwritten digits for zip code recognition in the postal service. This model largely introduced the convolutional neural network for the world. This architecture is considered as one of the simplest architectures. It has a total of 2 convolutional and 3 fully-connected layers (hence “5”). Most of the time, it is very common for the names of neural networks to be derived from the number of convolutional and fully connected layers that they have.

![Fig 2.6: Basic architecture of LeNet-5](image)

- **AlexNet**

AlexNet has a total of 8 layers (5 convolutional layers and 3 fully-connected layers). The general architecture of this model is quite similar to the LeNet-5, although this model is considerably larger. The AlexNet architecture mainly stacked a few more layers onto LeNet-5. The authors of this architectures pointed out that their architecture was “one of the largest convolutional neural networks during the publication on the subsets of ImageNet.” The novelty of this architecture, as pointed by the authors, is that it is the first to implement Rectified LinearUnits (ReLUs) as activation functions.
• **VGGNet**

The VGG network architecture, as one of the well-known architectures, offers a deeper yet simpler variant of the convolutional structures. At the time of its introduction, this model was considered to be very deep. It is currently the most preferred choice in the community for extracting features from images (*CNN Architectures*). The novelty with this architecture is that the researchers designed deeper networks (roughly twice as deep as AlexNet).

Among the architectures that are discussed above, we are going to build our own CNN model architecture following the VGGNet CNN architectures due to the following reasons:

- It is currently the most preferred choice in the community for extracting features from images (*CNN Architectures*).
- It has deep neural network architectures when compared to other CNN architectures which leads to improved performance (*Common architectures in CNN*).
We will implement our model in Keras using ImageNet weights provided by the library. The image dataset which we have prepared will be used for the training and testing of the model we are going to implement.

2.5. Related Works

2.5.1. Related Works Based on Clinical and Laboratory Results

Clinical scoring system for diagnosing AA is among the cheaper, faster and non-invasive (does not involve any medical instrument) diagnosis technique which can be used widely as a primary diagnosis tool for acute appendicitis (Y, 2014).

In (Alvarado, 1986), proposed a new clinical scoring system, which is called the Alvarado clinical scoring system, to diagnose acute appendicitis from the clinical and laboratory findings of acute appendicitis suspected patients. This diagnosis system mainly consists of clinical symptoms, and laboratory findings, and has shown good performances in suspected appendicitis patients for a long period of time. The researcher in their study developed a new acute appendicitis early diagnosis tool by considering the symptoms and laboratory results having a total of 10 points. The study of the researchers shows that, a clinical score of 7 or more is strongly predictive of the presence of acute appendicitis disease, whereas a score between 5 to 6 is suggestive of possible appendicitis, and score below 5 is indicative of other causes of abdominal pain. This method evaluated by Hamid Kariman et al (“Evaluation of the Alvarado score in acute abdominal pain”, March, 2014). And as pointed out by the researchers, the limitation which is encountered in Alvarado study is that, it has proven that an Alvarado score that is positive for appendicitis would suggests that the patient has a 93% chance of having appendicitis whereas a negative Alvarado score suggesting a 26% probability of having appendicitis. In (M. Lahaye, 2015), the researchers agreed that the clinical scoring system developed by Alvarado is a good rule-in test for acute appendicitis diagnosis, but this alone does not adequately rule out the acute appendicitis diagnosis. Even if the Alvarado clinical scoring system has gained good acceptence, it cannot be used these days because of that all clinical and laboratory based diagnosis are resulting in 15%-30% negative appendectomy rates worldwide (R. Jeffry,
As the Alvarado method is the first internationally accepted acute appendicitis diagnosis tool, it is used for a long period of time worldwide.

In (S. G. Prabhudesai, 2008), proposed a new artificial neural network (ANN) model for the diagnosis of acute appendicitis disease. The researchers used a back propagation algorithm in their study. The contribution of their work is that they achieved a sensitivity and specificity value of 100% and 97.2% which indicates that there are no false negative results at all. This technique is more reliable when compared to other previous studies in that it can minimize the negative appendectomy rate. The limitation of the researchers is that the proposed model uses the clinical and laboratory results only of the suspected patients which still does not allow alone to perform surgery for the removal of the appendix. Currently the model gains low acceptance because further image modality analysis is required.

In (Kim, 2015), proposed and designed an ANN model by considering nine (9) variables from clinical and laboratory results as the main factors for the appendicitis diagnosis. The new developed model by the researchers showed an improved performance for the diagnosis of appendicitis compared to the Alvarado clinical scoring system. The main limitation with this study is that it Only depends in clinical and laboratory data which alone does not suggest the level of severity of the disease. In addition to that, this method has similar drawbacks to the Alvarado method and gain low acceptance now a day because based on clinical examination, high number of negative appendectomy rates of have been reported (R. Jeffry, 1987). As it has been suggested in (A. Raja, 2010), the negative appendectomy rate can drop to as low as 2 % if imaging is added to the diagnostic work-up. So it is very essential to improve this work by using the imaging modality for diagnosis which helps to identify the details of the appendicitis for a better decision.

In (M. Tez, 2008), Tez and Tez proposed a neurofuzzy logic for diagnosis of acute appendicitis by considering clinical and laboratory results. In their work, they proved that the neuro fuzzy systems which they designed can incorporate data from many clinical and laboratory variables to provide better diagnostic accuracy in acute appendicitis. Here again even if the diagnostic accuracy is better than the others, it still depends on only clinical and laboratory variables which still leads to false positive results and which lacks acceptance.
In (G. Mostbeck, 2016), the study by Mostbeck et al. suggested the following key points in their work:

- Diagnosis which relies on imaging modality specifically Ultrasound (US) based diagnosis should be the first measure which should be applied for diagnosing acute appendicitis.
- The use of Ultrasound imaging modality for diagnosing AA disease will decrease the ionization when compared to X-ray and it is a good choice for children’s and pregnant women.
- If the application of US for AA diagnosis results in non-visualization of the appendix, clinical reassessment should be done and if diagnosis remains unclear, other imaging modalities such as MRI or CT may be used.

The researchers concluded in their work that for AA diagnosis, US should be the first-line imaging modality and it has excellent specificity both in the pediatric and adult patient populations.

In general, all the above clinical and laboratory-based AA diagnosis systems have been used for so long. But now a day, they are getting low acceptance as different studies in this area indicates that the diagnosis using these techniques are resulting in false appendectomy (R. Jeffry, 1987). So the diagnosis of AA should mainly depend on imaging modalities as since this method reduces the negative appendectomy rates from 30% to 2% if properly applied (A. Raja, 2010).

2.5.2. Related Works Based On Imaging Modalities

In (Y. Kester, 2019), the researchers proposed an Automated Diagnosis of Appendicitis from Clinical Notes which is taken from Emergency Department. The researchers developed a new deep neural network model, which is a combination of a convolutional neural network (CNN), a recurrent neural network (RNN), and a residual net-work that is able to predict the probability of appendicitis given a free-text emergency department (ED) note and additional structured information (e.g., lab test results). The contribution of the researcher’s work is that:
• The performance of the developed neural network architecture is promising and close to the performance of ED doctors.

• The designed model can learn important features, signs, and symptoms of patients from unstructured free-text ED notes, which will help doctors to make better decisions for their diagnosis.

Even if the designed model has good performance, it has also the limitation that the proposed model only focuses on ED notes which is the prior analysis for the diagnosis but not the only one because if the ED notes indicate that the patient is suspected of the acute appendicitis, further medications like US, MRI or CT results are needed. So this method encounters similar problems which are listed for clinical and laboratory based methods. And like other clinical and laboratory based techniques, the proposed model result alone cannot be used for emergency operation.

In (A. Walid, 2019), the researchers proposed an automatic approach for diagnosing the acute appendicitis disease from CT scan image. In this study, the exact location of the appendix is first inferred by a reinforcement-learning (RL) agent. Then, next the appendix is classified by using CNN and finally the classifier result is used by a Region of Low Entropy (RLE) to obtain the final probability of acute appendicitis diagnosis. The proposed method obtained better results which increases the acceptance of the method sensitivity and specificity of 0.91 and 0.926 respectively. However, the proposed method has encountered the following drawbacks: Since almost all acute appendicitis diagnosis is done using the US due to its many advantages over the others, the method uses CT image which can be done as a final test where US result for acute appendicitis is negative and the patient remains sick. So it only works for such cases (it is limited to a few cases), because of its side effects, again CT cannot be totally suggested for children’s and pregnant women’s. So the method still fails to work in such cases. In addition to that, the method only focuses on the structure of appendix while there are many parameters like local abscess around the appendix, fluid surrounding the appendix and perforation at the tip of the appendix which helps to make better decision.

In (R. Balu, 2011), proposed a new AA diagnosis system that can extract appendicitis well by calculating the distance between the two ends of the appendicitis using the Euclidian
distance measure. The contribution of this study is that the researchers used image data rather than clinical (numerical data) for their study. In addition to that, they are able to extract the appendicitis part well and finally able to obtain good classification accuracy in their study. The main drawback with this study is that it only depends on the diameter of the appendicitis while there are many other factors like thickness of the appendicitis which should be considered too, even the diameter variation between different ages, perforation around the appendicitis, Local abscess formation and other fluid around the appendicitis which should be considered too. Even if the results obtained from the experiment can have good acceptance, from the experimental results, it can be seen that there are true negative values (i.e.: implying that the appendicitis is normal although it is abnormal).

In (K. Baek, 2015), the researchers proposed an Intelligent Automatic Appendix Extraction Method from Ultrasonography Based on Fuzzy ART and Image Processing techniques. The developed model is mainly about the extraction of the appendicitis part correctly by using fuzzy ART algorithm. The contribution of the researchers is that their experimental result proves that the proposed method is highly accurate (successful in 38 out of 40 cases) in extracting appendix. However, the study only focuses on extraction of the appendicitis which does not involve the diagnosis.

2.6. Chapter Summary

In this chapter, the basic working principles of neural networks, specifically the convolutional neural network together with its well-known architectures is discussed in detail. In addition to that, different existing works relevant to this study were presented. Among the existing acute appendicitis diagnosis methods, the two main approaches namely diagnosis based on clinical and laboratory findings and diagnosis based on different imaging modalities are described. The existing relevant works were compared with one another, their contribution and the gap in their work is analyzed and presented well.
CHAPTER 3

METHODOLOGY

3.1. Overview
The methodology part covers the overall system architecture and proposed model architecture. In addition to that the detail of steps and tools that can be applied in this study will be briefly explained and the performance metrics that can be used to evaluate the proposed model will be presented well.

3.2. System Architecture
The proposed system architecture consists of image acquisition, pre-processing, feature extraction, image classification using proposed model, and performance evaluation of the proposed model. In the image acquisition part, the original image is given to the pre-processing phase. In the image pre-processing phase, the input image is pre-processed to increase its visibility by removing undesired distortions. In the feature extraction phase, the image features and labels can be extracted. The proposed model accepts the pre-processed image features and labels for training and validating the proposed model. The performance evaluation phase consists of testing the performance of the model using test data which has not been seen by the model. The detailed system architecture is is presented in Figure 3.1.
Fig 3. 1: Detailed system architecture
3.3. **Proposed Model Architecture**

The model which we have developed has the following specifications (layers). The basic components of the proposed layer consist of a convolutional layer, max-pooling layer, and output(dense) layer.

The proposed model has 5 convolutional layers:

- The first layer has 32-3x3 filter
- The second layer has 32-3x3 filter
- The third layer has 64-3x3 filter
- The fourth layer has 128-3x3 filter
- The fifth layer has 256-3x3 filter

In addition to these convolutional layers, there are five max-pooling layers each of size 2x2. At the end of the layers, there are two dense layers with the first layer having 64 outputs and the second layer having 1 output (which is the output layer)

The detailed architecture of the proposed model is presented below:
Fig 3.2: Proposed model architecture
3.4. Dataset Preparation

The dataset used for this study were collected from different internationally available repositories like (pediatrics, 2011) (abdomen and retroperitoneum, 2012) (appendicitis during pregnancy, 2013) (encyclopaedia) (Infant with perforated appendicitis, 2014). Our image data sets has been saved in a single image format (.JPEG file extenison) in order to solve an image classification problem.

For our experimental study, we have a single folder, containing a dataset for two classes (both positive and negative class). The folder contains a total of 4000 images among which 70% of the data can be used for training the model, 15% of the data for validating the model, and the remaining 15% for testing the classification performance of the model.

3.4.1. Data Augmentation

The overall model performance of any deep learning algorithms often depends with the amount of data available. But most of the time, there is scarcity of data set in medical field for applying deep learnings. When we train a large neural network with few training samples, we may face problems like overfitting. To avoid model overfitting and improve overall model performance, image augmentation will be applied to our input image data. In general, image augmentation is a well-known technique that can be used to artificially create new training data from existing training data. This technique is most widely used in medical image analysis sector as there is scarcity of image data in the sector. The augmentation follows different domain specific techniques to prepare different and new image samples from the available training data.

The transforms include a range of different operations from the field of image manipulation, such as shifts, flips, zooms, and much more. The need for data augmentation is to expand the training dataset with new, plausible examples. It means that, variations of the training data set that are likely to be seen by the model during training (Machine learning mastery).
3.5. Loading and Pre-processing Images

Image is a visual representation of real-world object. Image processing in computer vision context can be defined as a method to perform some operations on an image, in order to get an enhanced image or to extract some useful information from it. In general, the image processing technique is a process of image to image transformation where image is taken as input and the modified version of the image will be produced. Noise can be found in any type of image and it is perceived as degradation of quality. When we come to medical imaging modalities, speckle interference and low contrast are the main restrictions of ultrasound imaging. Speckle noise is classified under the category of multiplicative noise which can make it difficult to observe and interpret the ultrasound images for medical diagnosis (Lenrex, 2018).

Filtering is considered as one of the commonly used techniques used to reduce speckle noise. Low contrast in an image also leads to performance degradation in different aspects of image processing applications. In particular, it makes the edge detection more difficult, which may lead to a severe degradation to the overall processing quality (Carmina, 2016) (N.Bardun, 2012).

Ultrasound image and other imaging modalities such as CT and MRI have been widely used to assist medical diagnosis in contemporary medicine. However, the availability of noise on those images during the process of imaging, and data compression during the process of storage and transmission always interrupt the image quality, resulting in unreliable performance of the post-processing steps in the automated diagnosis systems such as medical image segmentation, feature extraction, and medical image classification. (C. Jianning, 2019).

The main objective of image pre-processing in any field of application specifically in medical diagnosis is to improve the image data quality by suppressing undesired distortions and also enhancing the required image features for further processing. The irrelevant data present in the image can be eliminated using the pre-processing technique. The pre-processing technique which is mainly applied on images before any image processing applications can eliminates the incomplete, noisy and inconsistent data from the image in the training and test phase. Since we do have already classified data (AA both positive and
negative images) set from the experts, we do not need image segmentation in image pre-processing steps.

• **Input Image**

The input image that can be given to our system is an appendicitis suspected patients Ultrasound image.

![Input image of AA +ve patient](image1.jpg) ![Input image of AA –ve patient](image2.jpg)

Fig 3.3: (a) Input image of AA +ve patient (b) Input image of AA –ve patient

There are different available image pre-processing techniques and few of the techniques which have been used for this study are discussed below.

• **Image Resizing**

Image resizing is considered as one of the important role in image processing technique, to enlarge and reduce the given image size in pixel format. Image interpolation can be divided into two different ways, they are image down-sampling and up-sampling which is necessary when resizing the data for matching either the specific communication channel or the output display (A.Schowengerdt, 2006).

An accurate resizing of image data in any image processing applications is an essential step, ranging from several consumer products to critical functions within the medical, security, defense and any other sectors.
• **Contrast Enhancement**

Images of good contrast are very essential in different image processing applications which helps both humans and machines to understand the behaviour of the image well. Contrast of any image can be improved by increasing the gray level shades among the objects in an image and background of an image. By means of different data transformation method, e.g., histogram equalization, wavelet transform etc., the contrast of an image can be enhanced. Since, histogram equalization method is one of the simplest and effective method, HE-based contrast enhancement is employed in many applications. However, some undesired Gray level variations are introduced in the image during the enhancement by the HE processes, those degrade the image quality. So, it is important to employ some procedure in the HE processes that can erase these variations (N.Bardun, 2012).

• **Noise Removal**

The mean filter and the Gaussian filter are typical linear filtering techniques that are effective and simple in smoothing speckle noise reduction. Moreover, the Gaussian filtering technique shows an excellent performance when compared to other techniques in terms of noise removal with a small variance; however, a blurring phenomenon appears in the edge areas. In our case, since the edge is not our focus area because we focus on the appendicitis which usually founds at the middle of our image data set, we will use the Gaussian filtering technique for the noise removal in our research work (A.Schowengerdt, 2006).

The Gaussian Filter which is one of the image noise removal filters which is similar to the mean filter however it involves a weighted average of the surrounding pixels and has a parameter sigma. The kernel represents a discrete approximation of a Gaussian distribution. Even if it is believed that the Gaussian filter blurs the edges of an image (like the mean filter) it does a better job of preserving edges than a similarly sized mean filter. It is also possible to specify the standard deviation for the x and y directions separately. If only one sigma value is specified during the image filtering process, then it is considered the sigma value for both the x and y directions (Towards data science).
Since we are using an image, which is a two-dimensional object, the Gaussian filter function for our images is given as:

\[ G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \]  

(3.1)

Where:

- \( x \) and \( y \) are the image pixel values
- \( \sigma \) is the standard deviation of the distribution

### 3.6. Feature Extraction

A feature of an image represents a piece of information of an image which helps us to solve the computational task related to certain application. For our image classification task, we first extract the image features and pass this feature to our model for training and validation. During the training and validation phase, features which are not relevant for our model will be dropped and only informative and non-redundant image features will be used.

### 3.7. Training and Classification

Training a model involves learning (determining) good values for our model which makes the model to perform the intended task with good performance. The type of learning algorithm which will be implemented in this study is a supervised learning where the function that maps an input to output based on example input-output pair will be determined.

Image classification in a simple term is the process of predicting the class of given image data. It involves extraction of all necessary information from an image and associating this information to its corresponding class labels.

The image classification task which will be implemented in our study is a supervised image classification where the model will be trained with input-output pair.

The classes in machine learning context are also called targets or labels and, in this study, the binary image classification task will be done where our input data for the proposed model has two classes namely the AA positive and AA negative classes. In our study, the
proposed model utilizes 70% of the total dataset for training and 15% of the total dataset for validating the model which helps the model to understand how the given input images relate to the classes they belong to.

3.8. Performance Evaluation Metrics

A model evaluation metric is one of the standard criterions by which the performance of a model is measured using different parameters. The different performance evaluation metrics of our proposed model were evaluated from the confusion matrix.

There are various ways to evaluate a classification model, and below are some of the most popular ones (A confusion matrix in machine learning).

3.8.1. Confusion Matrix

A confusion matrix which represented using table tells us the test prediction summary of our classification report. In confusion metrics, the number of correct and incorrect predictions are summarized with count values and these predictions are classified by each class used for both training and testing. It shows the ways in which our classification model is confused when it makes predictions. It gives us a detailed insight not only into the errors being made by a classifier but more importantly the types of errors that are being made during the predictions (A confusion matrix in machine learning).

We need to have the actual response class values and the predicted values to determine the confusion matrix. A confusion matrix is highly interpretative for classification problems and it can be used to estimate a number of other metrics.

![Confusion Matrix](image)

Fig 3.4: Confusion matrix for binary class
In the above confusion matrix, the term:

- **True positive**: the observation which was true is predicted true
- **True negative**: the observation which was negative is predicted negative
- **False Positive**: the observation which was negative is predicted as positive
- **False Negative**: the observation which was positive is predicted as negative

As it can be seen from the confusion matrix diagram above, the diagonal elements predict the correct predictions whereas the non-diagonal elements represent the samples which are miss-classified by the proposed model.

### 3.8.2. Classification Accuracy

The classification accuracy is the most popular used model evaluation metric for classification problems. The classification accuracy is the ratio of correct predictions of the model to the total number of predictions. It tells us the percentage of correct predictions made by the model.

Mathematically, it can be calculated as:

\[
Accuracy = \frac{\text{No of correct predictions}}{\text{Total No of predictions}}
\]

It can also be calculated from the confusion matrix obtained from the model. Following is the equation to calculate the accuracy using the confusion matrix:

\[
Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.2}
\]

### 3.8.3. Sensitivity/Recall

The recall can be defined as how sensitive the classifier is for detecting positive samples. A high recall value indicates that the class is correctly recognized (a small number of FN). It is also called a true positive rate. It can be represented mathematically as:

\[
Recall = \frac{\text{No of correct positive predictions}}{\text{Total Number of Actual positives}}
\]
From the confusion matrix, this recall can be calculated as:

\[
Recall = \frac{TP}{TP+FN} \quad (3.3)
\]

### 3.8.4. Specificity

Specificity is defined as the ratio of correctly classified negative predictions to the total number of negative predictions. This determines how specific the classifier model is in predicting negative samples.

It can be represented mathematically as:

\[
Specificity = \frac{No \ of \ correct \ Negative \ predictions}{Total \ Number \ of \ Negative \ predictions}
\]

So, our model specificity can be computed as:

\[
Specificity = \frac{TN}{TN+FP} \quad (3.4)
\]

### 3.8.5. Precision

Precision is defined as the ratio of correct predictions to the total number of correct predictions. It measures how precise the classifier is when predicting positive instances.

The precision of a model can be represented mathematically as:

\[
Precision = \frac{No \ of \ correct \ positive \ predictions}{Total \ Number \ of \ postive \ predictions}
\]

The precision of a model can be calculated from the confusion matrix as follows:

\[
Precision = \frac{TP}{TP+FP} \quad (3.5)
\]
3.8.6. F-measure

Depending on application, you may want to give higher priority to recall or precision. But there are many applications in which both recall and precision are important. Therefore, it is natural to think of a way to combine these two into a single metric. **One popular metric which combines precision and recall is called F1-score**, which is the harmonic mean of precision and recall defined as:

The F-measure of the model can be calculated as:

\[
F - measure = \frac{2 \cdot \text{Recall} \cdot \text{Precision}}{\text{Recall} + \text{Precision}}
\]  

(3.6)

3.9. Developing tools

The proposed system to diagnosis acute appendicitis from US image using CNN is implemented using the following description:

1. Programming languages and Integrated development environments
   (a) Python programming languages
   (b) Anaconda IDE
   (c) Matlab 2017a for image preprocessing
2. Libraries
   (a) Keras: which is an open-source neural-network library written in Python. It is capable of running on top of tensor flow, and other important libraries
3. Experimental setup

A desktop computer was used to conduct the experiment and addition virtual GPU from google colab is used. The specification for the desktop computer is given below.

<table>
<thead>
<tr>
<th>Processor</th>
<th>Intel core-i5</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAM</td>
<td>12GB</td>
</tr>
<tr>
<td>Operating system</td>
<td>Windows 10</td>
</tr>
<tr>
<td>HDD</td>
<td>1TB</td>
</tr>
</tbody>
</table>
3.10. Training The Model

For training our image classification model, we require:

- Training images
- Validation images
- Colaboratory IDE with virtual GPU
- Proposed CNN model
- Defining initial parameter like epoch, batch size and etc

3.11. Model parameters

**Epoch:** is a hyper parameter that is used to define the number times that the learning algorithm will work through the entire training dataset. In a deep learning context, One Epoch is when an entire dataset is passed forward and backward through the neural network only once. It determines when the neural network stops training.

The recommended value that can be used for any neural network application is an integer domain \([1, \infty)\).

In our case, we used to start our training using an early stopping method which stops the model when the validation accuracy stops to decrease for consecutive 3 epochs. Then our model was able to stop training at epoch 123 which increased the performance and decreased the loss of our model.

**Batch size:** is number of training samples present in a single batch. It is one of the very essential deep learning model parameters that defines the number of training samples to work through before updating the internal model parameters.

As we cannot pass the entire dataset to a neural network once, so we need to divide the dataset in to a number of batches, sets or parts. In our case, we used a batch size of 128 which resulted in a good performance of the model.

**Steps per epoch:** defines the number of batch iterations before a training epoch is considered finished. Most often, the steps per epoch is calculated as train data length//batch size.so in our case, steps per epoch is calculated using the above formula.
3.12. Chapter Summary

In this chapter, the hardware and software specifications that can be used in the study were presented. The techniques used to process images to increase the quality of an image were also described in a step wise manner. In addition to that the overall architecture of the proposed model is presented and each of the parameters used in the architecture is also described. The performance evaluation metrics are described in detail.
CHAPTER 4

RESULT AND DISCUSSION

4.1. Overview

In this chapter, experimental evaluation of the proposed model for automatic acute appendicitis diagnosis is presented in detail. The experiment result shows the realization of the proposed system architecture. The results obtained from pre-processing steps and results from the implementation of the model is presented clearly. The results obtained from the experiments were used to evaluate the performance of the proposed model.

4.2. Output of Image Pre-processing

The output image is an improve version of the input image with good data quality by which all undesired distortions are suppressed and also enhanced image with all the required image features for further processing.

Fig 4.1: Output image

(a) Original AA +ve Image

(b) Noise free AA +ve Image

(c) Contrast enhanced AA +ve Image
4.3. Summary of proposed model architecture

After creating the proposed model, we compiled it using Adam optimization algorithm which is the most well-known optimization algorithm. In addition to that, the loss function which is used for our model is binary-cross entropy as we are dealing with binary classification problem. Finally, we specified the metrics as accuracy in which we want to evaluate the model during training and validation. So the summary of the proposed model architecture is shown below:

Fig 4.2: Proposed model summary
4.4. Model performance for different learning rates

The model is trained for different learning rates for the same data (train test validate= (70%, 15% and 15%)) for 50 epochs. The learning rates are compared using the model validation accuracy.

Table 4. 1: Model performance on learning rates

<table>
<thead>
<tr>
<th>No</th>
<th>Learning rate</th>
<th>Optimizer</th>
<th>Validation Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.01</td>
<td>Adam</td>
<td>54.18</td>
</tr>
<tr>
<td>2</td>
<td>0.001</td>
<td>Adam</td>
<td>97.29</td>
</tr>
<tr>
<td>3</td>
<td>0.0001</td>
<td>Adam</td>
<td>88.45</td>
</tr>
</tbody>
</table>

4.5. Model performance for different train test split ratio

For this study, the image datasets which were collected from international repositories are divided into three parts namely training, validation and testing set. To select the training, validation and testing data ratios, the data is divided for 3 different ratios and all data are trained and validated for the learning rate of 0.001 (which is selected as good from the table above) and trained for the same 50 epochs. To select the best partitioning ratio, the model validation accuracy is measured and the result obtained is presented below.

Table 4. 2: Model performance for different data split

<table>
<thead>
<tr>
<th>No</th>
<th>Train, Validation and Test ratio</th>
<th>Learning rate</th>
<th>Validation Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>80%, 10%, 10%</td>
<td>0.001</td>
<td>95.92</td>
</tr>
<tr>
<td>2</td>
<td>70%, 15%, 15%</td>
<td>0.001</td>
<td>97.18</td>
</tr>
<tr>
<td>3</td>
<td>60%, 20%, 20%</td>
<td>0.001</td>
<td>96.37</td>
</tr>
</tbody>
</table>

From the above table, it is clear that train validation and test split ratio with good validation is the one with train, validation and test ratio of 70%, 15% and 15% respectively. So to train our model, we used this approach.
4.6. Implementation

The proposed model and other pre-trained models are all trained and validated based on the hyper parameters listed below in table.

<table>
<thead>
<tr>
<th>No</th>
<th>Learning rate</th>
<th>Optimizer</th>
<th>Batch size</th>
<th>Epoch</th>
<th>Data split ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.001</td>
<td>Adam</td>
<td>128</td>
<td>123</td>
<td>70%, 15%, 15%</td>
</tr>
</tbody>
</table>

4.6.1. Image Classification on Pre-Processed Data

Training and validation accuracy and training and validation loss of the proposed model is presented using figure below:

Fig 4.3: Model Accuracy on processed data
The proposed model which is trained for 70% of total data and validated on 15% of total data is tested for 15% (300 images for each class) of the total data and the model performance summary is computed from the below confusion matrix.

The confusion matrix result obtained from the test data from the trained model is shown below using table:

<table>
<thead>
<tr>
<th></th>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>296</td>
<td>2</td>
</tr>
<tr>
<td>Negative</td>
<td>3</td>
<td>299</td>
</tr>
</tbody>
</table>

The performance summary which is computed from the proposed model on pre-processed data is presented below using table.

<table>
<thead>
<tr>
<th>Accuracy</th>
<th>Loss</th>
<th>Recall/Sensitivity</th>
<th>Specificity</th>
<th>Precision</th>
<th>F_Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>99.16</strong></td>
<td><strong>2.133</strong></td>
<td><strong>98.99</strong></td>
<td><strong>99.33</strong></td>
<td><strong>99.32</strong></td>
<td><strong>99.15</strong></td>
</tr>
</tbody>
</table>
As it can be seen both from the confusion matrix and performance summary table, the results indicate that the model is able make correct predictions with prediction accuracy of 99.16%. There are also false predictions that indicates the model is classifying images in a false way.

4.6.2. Image Classification On Un-Processed Data

Training and validation accuracy and training and validation loss of the proposed model on unprocessed data is shown below using figure:

![Model Accuracy](Fig 4.5: Model Accuracy on un-processed data)

![Model Loss](Fig 4.6: Model Loss on un-processed data)
Table 4. 6: Confusion matrix on unprocessed data

<table>
<thead>
<tr>
<th></th>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>278</td>
<td>27</td>
</tr>
<tr>
<td>Negative</td>
<td>2</td>
<td>295</td>
</tr>
</tbody>
</table>

Table 4. 7: Performance summary on un-processed data

<table>
<thead>
<tr>
<th>Accuracy</th>
<th>Loss</th>
<th>Recall/Sensitivity</th>
<th>Specificity</th>
<th>Precision</th>
<th>F_Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>95.18</td>
<td>13.33</td>
<td>99.28</td>
<td>91.61</td>
<td>91.14</td>
<td>95.03</td>
</tr>
</tbody>
</table>

From the above results, it can be seen that the overall model performance is low on unprocessed data when compared to the processed one. The result tells us that an image pre-processing technique are very essential for medical images before further processing and increase the performance of the proposed model.

4.6.3. Image Classification on Pre-Processed Data with VGG19 pre-trained model

![Model Accuracy](image)

Fig 4. 7: Model Accuracy on VGG19 model for processed
Here, the dataset prepared for the proposed model is used to train, validate and test the model. Because we are applying this model for our problems, the output layer of the pre-trained VGG19 architecture is modified so that it works well for our problem.

![Model Loss Graph]

Fig 4.8: Model Loss on VGG19 model for processed

<table>
<thead>
<tr>
<th>Table 4.8: Model test loss and test accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test loss</td>
</tr>
<tr>
<td>Test accuracy</td>
</tr>
</tbody>
</table>

From the table above, we can see that the model has good accuracy while training and validating and the model loss is very high. While it has good classification accuracy when compared to our model, there is a significantly high loss for model in VGG19. In general, the pre-trained models are normally trained for natural images which has a big difference compared to medical image in many aspects.

- Almost all pre-trained models are trained for a very large number of images and the image scarcity in medical field may also affects the performance.
- The model is learning very deep features of images during training and validation while it is not doing the same during testing.
4.6.4. Image Classification On Pre-Processed Data with InceptionNetV3 pre-trained model

Here, the dataset prepared for the proposed model is used to train, validate and test the model. Because we are applying this model for our problems, the output layer of the pre-trained VGG19 architecture is modified so that it works well for our problem.

![Model Accuracy on InceptionNetV3 model for processed data](image)

**Fig 4.9:** Model Accuracy on InceptionNetV3 model for processed data

![Model Loss on InceptionNetV3 model for processed data](image)

**Fig 4.10:** Model Loss on InceptionNetV3 model for processed data
As it can be seen from both the figure and performance summary of the model, the result shows there is very high model loss for InceptionNetV3 pre trained model when compared to our model performance.

In both VGG19 and InceptionNet pre-trained models, the results show that pre-trained models perform poor in medical image classification task and it may not help us much as it is mainly trained for natural images. So in our case, our model which is developed from scratch works well than pre-trained ones.

So the transfer learning in our case is not a better choice than training from scratch as it learns very high-level features while training and fails to validate and test on unseen data. We can conclude that transfer learning is not performing well in medical image classification tasks because our model which is designed from the scratch is performing better.

4.7. Answers to Research Questions

At the beginning of this study, we mentioned three research questions in chapter one. Now we will try to answers the research questions based on the results obtained from the experiment.

Research question 1: To what extent do deep learning algorithms can diagnosis acute appendicitis from US image?

As it can be seen from the experimental results, deep learning algorithms specifically the convolutional neural network performed an excellent level in our case by achieving an accuracy of 99.16 for pre-processed data and 95.18 for unprocessed data respectively.
Research question 2: By what percent do image pre-processing can increase the performance of the system?

It is clear from the comparison of the model on both pre-processed and unprocessed data that our proposed model has a significant change with 4.013% increase for pre-processed data.

4.8. Chapter Summary

In this chapter, the results obtained from the experiment of the proposed model is explained in detail and the performance of the model is evaluated using different evaluation metrics. In addition to that, the results obtained from the experiments on both processed and unprocessed data is compared. In general, the results obtained from the experiments on pre-processed data has obtained excellent results and our proposed model is outperforming other pre-trained models tested in this study.
CHAPTER 5

CONCLUSION AND FUTURE WORK

5.1. Conclusion

Among different working domains available worldwide, Medical industry is one of the well-known domains where predictions are very important. This research work encourages the use of artificial intelligence (AI) as technological support for decision-making in medical field for diagnosing AA disease. The motivation behind the development of the disease diagnosis model is not because the human intelligence would not be enough, but because the mechanisms of artificial intelligence have several benefits that are suitable for this field.

Acute appendicitis is one of the most common abdominal pain that leads to emergency surgery. Currently, US has gained high acceptance for diagnosis of patients with severe abdominal pain because it can have no any radiation, and is easily available. It is obvious that manual analysis of US image needs an expert to interpret the result and in addition to that, it is tedious and time-consuming task. To avoid such like problems, it is necessary to have a model that can do the above task with minimum amount of time, labor and with no need for an expert.

The work presented in this thesis is all about the development of an intelligent model that can help physicians in decision making by detecting acute appendicitis from ultrasound images. In this work, image pre-processing techniques were applied in order to improve the quality of ultrasound images. The proposed model is tested for both pre-processed and unprocessed dataset and the results show that the proposed model achieved an accuracy of 99.16 for pre-processed data and 95.18 for unprocessed data respectively. There is also a significant change in model loss in both cases. As it can be seen from the experimental results, the proposed system achieved good performance on pre-processed images for the model.

This study proposes an intelligent appendicitis detection system from ultrasound images. The ultrasound images which can be used for the proposed system are pre-processed using
different image pre-processing techniques. The image pre-processing techniques provided us an image with high quality which can be used for further diagnosis system. The proposed system has many advantages like high model accuracy and low loss when compared to previous existing systems. The study contributes for the contemporary medicine in different aspects like simplifying the physicians’ work, saving their time and energy (which otherwise would be wasted on too many things they have to do). This way, the physicians can focus on more important activities. On the other hand, an automated system can detect invisible features which cannot be detected by humans.

Finally, it can be concluded that it is difficult to point out that our proposed system can replace humans. But from our study, we are sure that automated systems can be of much help for medical professionals in the decision making process.

5.2. Contribution

This research study has a lot of contribution to the scientific world at large. The contributions include:

- The proposed model can detect AA with good classification accuracy.
- The study helps the medical world by making the diagnosis system simple which supports clinicians on decision making.
- The new proposed system can solve the problems with the existing systems and previous studies by considering other parameters like thickness of the appendicitis, perforation of the appendix, fluid surrounding the appendix which are not yet considered.
- The model will improve the overall management outcome of the patient by automating the overall appendicitis diagnosis system.
5.3. Future Work

As machine learning is increasingly used in real-world decision processes, the necessity for intelligent systems that can help humans in decision making in all aspects of life will continue to grow. There are several possibilities of future work related to this thesis work that can increase the flexibility and performance of the proposed work. An extension of this work could be, since our work is limited to the diagnosis of AA from US images only, it can be further extended to diagnosis an acute appendicitis by combining both abdominal US images and structured laboratory results.
REFERENCES


55


N. Bardun. (2012, August). Contrast enhancement with the noise removal by a discriminative filtering process”.


APPENDIX

I. Training and Validation Phase

# import the necessary packages
from keras.layers import Activation, Dropout, Flatten, Dense
from keras.preprocessing.image import ImageDataGenerator
from sklearn.model_selection import train_test_split
from keras.preprocessing.image import img_to_array
from keras.layers import Conv2D, MaxPooling2D
from keras.callbacks import EarlyStopping
from sklearn.metrics import confusion_matrix
from sklearn.metrics import precision_score
from sklearn.metrics import accuracy_score
from sklearn.metrics import recall_score
from keras.utils import to_categorical
from sklearn.metrics import f1_score
from keras.models import Sequential
from keras.optimizers import Adam
import matplotlib.pyplot as plt
from keras import backend as K
from imutils import paths
import numpy as np
import argparse
import random
import cv2
import os

#Initial paraamaeters
IMG_SIZE=200
EPOCHS =123
BS=128
input_shape = (IMG_SIZE, IMG_SIZE, 3)# dimensions of our images.
opt=Adam(lr=0.001)

#partition the dataset in to three parts
#for training, validation and testing, we use 70%, 15% and 15% respectively

train_ratio = 0.70
validation_ratio = 0.15
test_ratio = 0.15

#Giving path to the dataset directory with its classes
image_dir="/content/gdrive/My Drive/Research_Data/Model_Data"
CATEGORIES=['Negative','Posetive']

for category in CATEGORIES:
    path=os.path.join(image_dir,category) #A path to either a negative or posetive set
    for img in os.listdir(path):
        img_array=cv2.imread(os.path.join(path,img))
        #data.append(img_array)
        #plt.imshow(img_array,cmap="gray")
        #plt.show()
        break
    break
new_array=cv2.resize(img_array, (IMG_SIZE, IMG_SIZE))

model_data=[]
def create_model_data():
    for category in CATEGORIES:
        path=os.path.join(image_dir, category)  # A path to either a negative or positive set
        class_num=CATEGORIES.index(category)
        for img in os.listdir(path):
            try:
                img_array=cv2.imread(os.path.join(path, img))
                new_array=cv2.resize(img_array, (IMG_SIZE, IMG_SIZE))
            except Exception as e:
                pass
            model_data.append([new_array, class_num])
create_model_data()
# print(len(model_data))

Features=[]  # Features
labels=[]  # label

for features, label in model_data:
    Features.append(features)
    labels.append(label)
Features=np.array(Features).reshape(-1, IMG_SIZE, IMG_SIZE, 3)
#now we assign 70% of the total data to training set
#and the remaining 30% for test first
x_train, x_test, y_train, y_test = train_test_split(Features, labels, test_size=(1 - train_ratio))

# now we divide the 30% test data in to two equal partitions
#namely the test set and the validation set
#now the train test and validation data will be 70%, 15% and 15% respectively
x_val, x_test, y_val, y_test = train_test_split(x_test, y_test, test_size=test_ratio/(test_ratio + validation_ratio))

# print(x_train, x_val, x_test)

# We need to rescale the pixel values in range 0 - 1 inclusive.
# For training set, we will do it later
x_test=x_test/255.

# Defining the model
my_model = Sequential()
my_model.add(Conv2D(32, (3, 3), input_shape=input_shape))
my_model.add(Conv2D(32, (3, 3)))
my_model.add(Activation('relu'))
my_model.add(MaxPooling2D(pool_size=(2, 2)))
my_model.add(Dropout(0.2))
my_model.add(MaxPooling2D(pool_size=(2, 2)))
my_model.add(Dropout(0.2))

my_model.add(Conv2D(128, (3, 3)))
my_model.add(Activation('relu'))
my_model.add(MaxPooling2D(pool_size=(2, 2)))
my_model.add(Dropout(0.2))

my_model.add(Conv2D(256, (3, 3)))
my_model.add(Dropout(0.2))

my_model.add(Flatten())
my_model.add(Dense(64))
my_model.add(Activation('relu'))
my_model.add(Dropout(0.2))

my_model.compile(loss='binary_crossentropy',
        optimizer=opt,
        metrics=['accuracy'])

# this is the augmentation configuration we will use for training
train_datagen = ImageDataGenerator(
    rescale=1. / 255,)
# this is the augmentation configuration we will use for validation
valid_datagen=ImageDataGenerator(
    rescale=1. / 255,)
train_set = train_datagen.flow(x_train, y_train, batch_size=BS)
valid_set = valid_datagen.flow(x_val, y_val, batch_size=BS)

# training the model
model_history = my_model.fit_generator(train_set,
    validation_data=valid_set, steps_per_epoch=len(x_train) // BS,
    epochs=EPOCHS, verbose=1)

# Early stopping
Early_stop = EarlyStopping(monitor='val_loss', mode='min', verbose=1, patience=3)

# Storing the details from history
model_accuracy = model_history.history['accuracy']
model_val_acc = model_history.history['val_accuracy']
model_loss = model_history.history['loss']
model_val_loss = model_history.history['val_loss']

epochs = range(1, len(model_accuracy)+1)
# plotting the model training and validation accuracy
plt.plot(epochs, model_accuracy, 'b', label='Training Accuracy')
plt.plot(epochs, model_val_acc, 'r', label='Validation Accuracy')
plt.title('Model Accuracy')
plt.xlabel('Epoch')
plt.ylabel('Accuracy')
plt.legend()
plt.figure()
#plotting the model training and validation loss

plt.plot(epochs, model_loss, 'b', label='Training Loss')
plt.plot(epochs, model_val_loss, 'r', label='Validation Loss')
plt.title('Model Loss')
plt.xlabel('Epoch')
plt.ylabel('Loss')
plt.legend()
plt.show

# save model and architecture to single file
my_model.save('model_pro_final.h5')
print("Model successfully saved to disk")

II. Testing phase code

# import the necessary packages
from keras.layers import Activation, Dropout, Flatten, Dense
from keras.preprocessing.image import ImageDataGenerator
from sklearn.model_selection import train_test_split
from keras.preprocessing.image import img_to_array
from keras.layers import Conv2D, MaxPooling2D
from keras.callbacks import EarlyStopping
from sklearn.metrics import confusion_matrix
from sklearn.metrics import f1_score
from keras.models import Sequential
from keras.optimizers import Adam
from keras.optimizers import RMSprop
import matplotlib.pyplot as plt
from keras import backend as K
from imutils import paths
import numpy as np
import argparse
import random
import cv2
import os

# dimensions of our images.
img_width, img_height = 200, 200
EPOCHS = 123
BS = 128
input_shape = (img_width, img_height, 3)
opt = Adam(lr=0.001)

#Giving path to the dataset directory with its classes
image_dir = "/content/gdrive/My Drive/Research_Data/Model_Data"
CATEGORIES = ["Negative", "Posetive"]

for category in CATEGORIES:
    path = os.path.join(image_dir, category) # A path to either a negative or posetive set
    for img in os.listdir(path):
        img_array = cv2.imread(os.path.join(path, img))
        #data.append(img_array)
        #plt.imshow(img_array, cmap="gray")
        #plt.show()
        break
break
IMG_SIZE=200
new_array=cv2.resize(img_array,(IMG_SIZE,IMG_SIZE))

model_data=[
def create_model_data():
    for category in CATEGORIES:
        path=os.path.join(image_dir,category)  # A path to either a negative or positive set
        class_num=CATEGORIES.index(category)
        for img in os.listdir(path):
            try:
                img_array=cv2.imread(os.path.join(path,img))
                new_array=cv2.resize(img_array,(IMG_SIZE,IMG_SIZE))
            except Exception as e:
                pass

create_model_data()
# print(len(model_data))

Features=[]  # Features
labels=[]  # label

# partition the dataset into three parts
# for training, validation and testing, we use 70%, 15% and 15% respectively

train_ratio = 0.70
validation_ratio = 0.15
test_ratio = 0.15
# now we assign 70% of the total data to training set
# and the remaining 30% for test first
x_train, x_test, y_train, y_test = train_test_split(Features, labels, test_size=(1 - train_ratio))

# now we divide the 30% test data in to two equal partitions

#print(x_train, x_val, x_test)

# We need to rescale the pixel values in range 0 - 1 inclusive.
# For training set, we will do it later
x_test=x_test/255.

# Loading saved model
from keras.models import load_model
test_model = load_model('model_pro_final.h5')
#print("Model successfully loaded")

# Now the next step will be evaluating the performance of the model using test data

# predict probabilities for test set
test_prob = test_model.predict(x_test, verbose=0)
# predict classes for test set
clas_prediction = test_model.predict_classes(x_test, verbose=0)

# reduce to 1d array
test_prob = test_prob[:, 0]
clas_prediction = clas_prediction[:, 0]
# confusion matrix
confusion_matrix = confusion_matrix(y_test, clas_prediction)
print(confusion_matrix)

# predict probabilities for test set
y=x_test[100].reshape(1,200,200,3)
test_prob_1 = test_model.predict(y, verbose=0)
clas_prediction_1 = test_model.predict_classes(y, verbose=0)

# reduce to 1d array
test_prob_1 = test_prob_1[:, 0]
clas_prediction = clas_prediction_1[:, 0]
# print(clas_prediction_1)
# print(test_prob_1)
# plt.imshow(x_test[200])
probab=test_model.predict_proba(y, verbose=0)
print(probab)

# single test
from keras.preprocessing import image
image_shape=(200,200,3)
new_image_1= image.load_img('/content/gdrive/My Drive/Research_Data/Test_Images/Negative_1.jpg',target_size =image_shape)
new_image = image.img_to_array(new_image_1)
# print(new_image)
new_image = np.expand_dims(new_image, axis = 0)
new_image=new_image/255
new_image=new_image.reshape(1,200,200,3)
test_im=test_model.predict_proba(new_image, verbose=0)
clas_prediction_2 = test_model.predict_classes(new_image, verbose=0)

# reduce to 1d array
test_prob_2 = test_im[:, 0]
clas_prediction_2 = clas_prediction_2[:, 0]

print(test_prob_2)
result=test_prob_2

#Making predictions
if (result==1):
    prediction = 'unknow'
    prediction='AA Posetive'
    prediction = 'AA Negative'

print(prediction)

III. Codes for image resizing

""
Created on Tue Feb 25 07:52:49 2020

@author: Mechal
""

from PIL import Image
import os, sys

path = "C:/Users/Mechal/Desktop/Data_Final/step_3_noise/Posetive/"
dirs = os.listdir( path )

def resize():
    image_no=1
    for item in dirs:
        if os.path.isfile(path+item):
            im = Image.open(path+item)
            f, e = os.path.splitext(path+item)
            Resized = im.resize((200, 200), Image.ANTIALIAS)
            name = 'C:/Users/Mechal/Desktop/Data_Final/step_4_resiz/Posetive/Image_' + str(image_no)+'.jpg'
            Resized.save(name, 'JPEG')
            image_no += 1

    resize()
    print('Done!')

IV. Contrast Enhancement code

%Open folder selection dialog box, for selecting input and output folders.
indir = uigetdir(cd, 'Select input folder');
outdir = uigetdir(cd, 'Select output folder');
directory = dir([indir, '\', '*.jpg']);

for i = 1 : length(directory)
    filename = directory(i).name;
    orig_image = imread([indir, '\', filename]);

    img = histeq(orig_image);
    %Save gray image to outdir (keep original name).
imwrite(img, [outdir, '\', filename]);

end

V. Noise Removal code

%Open folder selection dialog box, for selecting input and output folders.
indir = uigetdir(cd, 'Select input folder');
outdir = uigetdir(cd, 'Select output folder');
directory = dir([indir, '\', '*.jpg']);

for i = 1 : length(directory)
    filename = directory(i).name;
    img = imread([indir, '\', filename]);
    noise_free = imgaussfilt(img);
    %Save gray image to outdir (keep original name).
    imwrite(noise_free, [outdir, '\', filename]);
end