DEEP LEARNING BASED CERVICAL CANCER DISEASE DETECTION AND CLASSIFICATION MODEL

A Thesis Presented

By

Nunu Gebeyehu

To

The Faculty of Informatics

Of

St. Mary's University

In Partial Fulfillment of the Requirements
For the Degree of Master of Science

In
Computer Science

January, 2023

Acceptance

DEEP LEARNING BASED CERVICAL CANCER DISEASE DETECTION AND CLASSIFICATION MODEL

By

NUNU GEBEYEHU

Accepted by the Faculty of Informatics, St. Mary's University, in partial fulfillment of the requirements for the degree of Master of Science in Computer Science

Thesis Examination Committee:

____

Internal Examiner

____

External Examiner

____

Dean, Faculty of Informatics

January 2023

Declaration

I hereby declare that this thesis entitled "**Developing Deep-Learning Based Cervical Cancer Disease Detection and Classification Model**" was prepared by me, with the guidance of my advisor. The work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted, in whole or in part, for any other degree or professional qualification.

Nunu Gebeyehu

Signature

Addis Ababa

Ethiopia

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

Alembante Mulu (PhD)

Signature

Addis Ababa

Ethiopia

January 2023

Certification Page

This is to certify that the thesis prepared by Mrs. Nunu Gebeyehu entitled "**Developing Deep Learning-Based Cervical Cancer Disease Detection and Classification Model**", submitted as partial fulfillment for the Degree of Master of Science complies with the regulations of the University, and meets the accepted standards with respect to originality, content, and quality.

Signed by Examining Board:

External Examiner: Signature, Date:

____

Internal Examiner: Signature, Date:

____

Chair Person: Signature, Date:

____

DGC Chair Person: Signature, Date:

____

January 2023

Abstract

Cervical cancer is the second most common and second most deadly cancer in Ethiopia. The disease's incidence and prevalence are increasing over time due to population growth and aging, as well as an increase in the prevalence of well-established risk factors. Cervical cancer knowledge and awareness among Ethiopian women is quite low. It is the most deadly disease caused by the uncontrolled growth of body cells, accounting for approximately 9.6 million deaths each year in world. In women, abnormal cell growth can affect various body organs such as the breast and the cervix. 85-90 percent of the fatality rate of cervical cancer occur in low and middle-income countries due to a lack of public awareness about the disease's causes and consequences. As a result, it is necessary to create a cervical cancer detection and classification model using deep learning techniques to assist experts. Sample of cervical cancer images were taken from Bethazeta Hospital in Addis Ababa, Ethiopia and some of data was addedfrom public dataset. It is proposed to detect and classify cervical cancer using deep learning model. The proposed approach has two main phases. In first phase the designed model is trained and tested by collected dataset and the data is classified using different neural network. Finally, the deep learning model that can detect and classify the given image in to Type_1, Type_2 and Type_3 is done. The dataset contains 2085 original cervical cancer images. From this, 80% of the images are used for training and the rest for testing the model. During training, data augmentation technique is used to generate more images to fit the proposed model using by Keras libraries. The Convolutional Neural Network and Hybrid of Convolutional NeuralNetwork and Long Short Term Memory model can successfully detect and classify the given image with an accuracy of 99.04% and 98.72% respectively.

Keyword: Deep Learning, Cervical Cancer, Convolutional Neural Network, Long Short TermMemory, Classification, Detection.

Acknowledgments

First and foremost thanks and respect be to the Almighty God.

I would like to thank my Advisor: Dr.Alembante Mulu. This thesis would not have existed without my advisor's continual direction, support, unrivaled expertise, and encouragement.

Besides my advisor, I would like to thank Dr.Bekele Bedede (Physician), Mr. Degefa Guta(Radiologist) at Bethazeta Hospital, who helped and supported me during the data collection and preparation stage of the experiment.

Lastly, I would want to express my heartfelt gratitude to my beloved Husband, and my lovely son, for unwavering love and support, for all of the late nights and early mornings, and for keeping me sane over the past several months.

Table of Contents

Abbreviations and Acronyms

ANN	Artificial Neural Network
CNN	Convolutional Neural Network
LSTM	Long Short Term Memory
ILSVRC	ImageNet Large-Scale Visual Recognition Competition
RNN	Recurrent Neural Network
NN	Neural Network
CCDCM	Cervical cancer Detection and Classification Model
ML	Machine Learning
DL	Deep Learning
SVM	Support vector machines
SVR	Support vector Regressions
RGB	Red Green Blue
GPUs	Graphics Processing Units
HOT	Histogram of Template
GAP	Global Average Pooling
GUI	Graphical User Interface
ReLU	Rectified Linear Unit
HIS	Hue, Saturation, Value
VGG	Visual Geometry Group
FC	Fully Connected

List of Tables

List of Figures

CHAPTER ONE

1. INTRODUCTION

1.1. Background

Cancer arises in the body when the cells of a specific organ start to grow abnormally. It is the most deadly disease caused by the uncontrolled growth of body cells, accounting for approximately 9.6 million deaths each year [1]. In women, abnormal cell growth can affect various body organs such as the breast and the cervix. Cervical cancer is a type of cancer that affects a woman's cervix. The cervix is the uterus's neck-shaped passage at the bottom. According to the most recent world cancer statistics, cervical cancer is the fourth most common cancer worldwide (with an estimated 570,000 new cases each year) and the second most common cancer among women living in developing and low-income countries, with an estimated annual death rate of 311,000 women, with 85-90 percent of these fatality rate occurring in low- and middle-income countries due to a lack of public awareness about the disease's causes and consequences [2].

Cervical cancer is linked to pregnancy and first intercourse at an early age, having sex with multiple partners, a weakened immune system, smoking, the use of oral contraceptives, and poor menstrual hygiene. Abnormal vaginal bleeding, vaginal discharge, and moderate pain during sexual intercourse are all common symptoms of cervical cancer. Cervical cancer may be treated if caught early and treated properly. Even if cervical cancer cannot be cured, it is often possible to delay its progression, extend life expectancy, and reduce any related symptoms including discomfort and vaginal bleeding which is known as palliative care.

There are two types of tests to know the occurrence of cervical cancer which are called the Pap test and HPV test [3]. The Pap test is sometimes called Pap smear that looks for pre-cancers, cell changes on the cervix that might become cervical cancer if they are not treated appropriately. The HPV test looks for the virus called human papillomavirus that can cause these cell changes.

Cervical cancer can be prevented and cured by removing affected tissues in early stages. So early prediction is needed to reduce an impact of cervical cancer spatially, in developing countries like Ethiopia. There are lack of expert in Ethiopia so applying different technologies which support in prediction is necessarily essential. Deep learning applications is a form of machine learning that teaches computers to do things that humans do naturally. Algorithms for machine learning have the ability to be extremely useful. Invested heavily in all aspects of medicine, from drug development to clinical decision-making, dramatically changing the landscape. So employing deep learning approach can bring significant changes in diagnosing cervical cancer at early stages. In this study convolutional neural network which is deep network will be applied to cervical cancer image obtained from medical device and classified as normal or abnormal.

1.2. Motivations of the Study

Most females in Africa mainly in sub-Saharan countries are affected by cervical cancer by the absence of early screening to the cancer. This is due to lack of radiologist that have the capability of testing for it. In Ethiopia also there is a lack radiologist. The physician are in high burden to support the patient with different cancer. Also there are many hospitals when it is compared to population ratio. So there is no suitable environment as other developed countries for checking their cervix frequently. Most women in Ethiopia are tested when the symptoms of cervical cancer is diagnosed manually. But this cancer treated properly if it is detected at early stage. This research aim to develop an automatic model that has the capability of detecting and classifying the cervical cancer type.

1.3. Statements of the Problem

Cervical cancer is the first most common cancer in women in sub-Saharan Africa followed by breast cancer. In Ethiopia, the incidence of cervical cancer is high i.e. 35.9 per 100,000 women [4]. Low level of awareness, lack of effective screening programs, overshadowed by other health priorities (such as acquired immune deficiency syndrome, tuberculosis and malaria) and insufficient attention to women’s health are the possible factors for the observed higher incidence rate of cervical cancers in the country.

Cervical cancer is most often diagnosed in women between the ages of 35 and 44 [4], with a median age of 50 at diagnosis. It rarely occurs in women under the age of 20. Many older women are unaware that the risk of cervical cancer remains the same as they get older. The 5-year survival rate for all people with cervical cancer is 66% [4]. Survival rates, on the other hand, can vary depending on factors such as race, ethnicity, and age. The 5-year survival rate for white women is 71 percent. The 5-year survival rate for Black women is 58 percent. This is also another issue that will be solved in Ethiopia by preparing a dataset collected from hospitals in Ethiopia.

Everyone with a cervix between the ages of 25 and 65 should be screened for cervical cancer every 5 years with an HPV test alone, according to the American Cancer Society. If HPV testing is not available, people can be screened every 5 years with an HPV/Pap co-test or every 3 years with a Pap test [4]. Nowadays in Ethiopia it is difficult to conduct both Pap test and HPV test due to lack of experts and device. So the proposed model can support the lack of expert in diagnosing and detecting cervical cancer at early stage.

The development of cervical cancer is usually slow and preceded by abnormalities in the cervix (dysplasia). However, the absence of early stage symptoms might cause carelessness in prevention. Additionally, in developing countries, there is a lack of resources, and patients usually have poor adherence to routine screening due to low problem awareness. Also as deep learning approaches are applied and detected at early stage it has its own impact in creating awareness about cancer. So this research is aiming to develop an automated cervical cancer detection and classification in deep learning technique. The CNN approach in deep learning is a state of artifact that can be applied in disease classification which has to be emphasized in reducing a problem in medical sector of Ethiopia.

1.4. Research Questions

This study addressed the following research questions and proposed solution

Q1. How deep learning algorithm and computer vision technique are used to detect and classify cervical cancer?

Q2. How dataset of cervical cancer could be prepared and performance of cervical cancer disease detection and classification model measured?

Q3. To what extent the proposed approach perform in detecting and classifying cervical cancer disease?

Q4. How to evaluate cervical cancer using evaluation parameter?

1.5. Objectives

1.5.1. General Objective

The general objective of this thesis is to develop deep-learning based cervical cancer Detection and Classification model by using CNN approach.

1.5.2. Specific Objectives

The following specific objectives will be accomplished to achieve the general objective of the study.

- To review related literature on the domain of the research,
- To prepare appropriate image dataset for training and testing the model,
- To develop a model for cervical cancer diseases detection and classification,
- To conduct a comparative analysis of both models,
- To evaluate the proposed model by using testing dataset.

1.6. Significance of the Study

The aim of this research is to use deep learning techniques to diagnose and identify cervical cancer. One of the machine vision technology that can solve the problem in health is deep learning. Supporting a physician or experts geographically in Ethiopia. Early cancer detection and prediction are critical in speeding up the diagnosis process and thus increasing life expectancy. If deep learning is used in this field, there are numerous benefits.

- Early cervical cancer disease detection
- Reduce abnormal vaginal bleeding if detected early
- Customizing a moderate model for Ethiopia

- Support an expert by AI
- Reduce expense of patients
- Increase survival rate of patients

1.7. Scope and Limitation of the Work

The thesis focused on using deep learning techniques to establish a cervical cancer disease detection and classification model. A new model will be used to detect and classify cervical cancer disease to their stage types. Images obtained from hospitals will be used to create data set for model. The dataset are image data and the classification of cancer type is based on Colposcopy cervical image. The limitation of this research is not consider other cancer type like breast cancer, brain cancer and so on is out of the scope of the study. This research work can not recommend the treatment of cervical cancer.

1.8. Contribution of the Study

I expect this thesis work would provide different benefits to the target populations in the domain area identified. The first contribution of this thesis for the research community and the whole population is to develop deep learning model CNN and Hybrid of CNN and LSTM to correctly detect and classify the cervical cancer in to Type-1, Type-2, Type-3 by using images that are taken from Bethazeta Hospital in Addis Ababa, Ethiopia and some of data was added from public dataset. The second main contribution of this thesis is the provision of well-organized and managed dataset of cervical cancer. To accomplish these, we have conducted several experiments by using different proposed models. Moreover, this research can be used as a reference for future researchers in the domain area.

1.9. Research Methodology and Design

A research methodology is a set of procedures, techniques, tools, and documentation aids that will assist the researcher in carrying out and completing the research work. The arrangement of conditions for data collection and analysis in order to address the research problem is known as research design [5]. The goal of this study was to develop a model for detecting and categorizing cervical cancer diseases. This is an experimental study. This includes conducting a literature review, preparing data sets, implementing procedures, tools, and evaluating performance.

1.10. Organization of the Thesis

This thesis report is organized into five different chapters. The first chapter presents a preliminary introduction to this study. It offers the general structure included in this study. It thus provides enough background information to help the reader understand the reason behind the study and what the researcher plans to accomplish by carrying out the research. The chapter provides an overview of the whole study. Chapter two of the study provides an explanation to the cervical cancer which will be selected for demonstrating the proposed methodology, deep learning and their type that are used in this research and present reviews of previous work related to the study topic with specific reference to the research objectives. It presents summaries from books, journals and collected works that are helpful in accomplishing this work and reflecting key conclusions and recommendations. Chapter three gives a details explanation of the datasets that are used in this research, the hyper-parameters range for the deep learning algorithms, the features that are extracted and used in this thesis. Chapter four presents the research results and a detailed analysis obtained through the methodology presented in chapter three. The last chapter, chapter five, presents a summary of results and draws conclusions from the study for users of the research and provides the future work for this study.

CHAPTER TWO

2. LITERATURE REVIEW AND RELATED WORKS

2.1. Overview

This chapter mainly focuses on the literature related to cervical cancer detection. It contains the background information and review of literature that are related to the domain of the thesis. The occurrence of cervical cancer, type and levels of cervical cancer are discussed. And also the data related to cancer are stated briefly. Technology available which has a contribution in detection of cervical cancer such as digital image processing, Machine Learning, Computer vision and deep learning concepts are illustrated well. Finally, a summary of related work is included in this chapter.

2.2. Cervical cancer overview

Cervical cancer is a form of cancer that develops in the cells of the cervix, which connects the uterus to the vaginal canal. Most cervical cancer is caused by different strains of the human papillomavirus (HPV), a sexually transmitted infection [6]. When the body is exposed to HPV, and the immune system usually stops the virus from causing harm. However, in a small number of people, the virus can live for years, contributing to the transformation of some cervical cells into cancer cells. You can lower your chances of acquiring cervical cancer by having screening tests and getting an HPV vaccine. Squamous cell tumors account for the majority of cervical malignancies (80–90%) [6]. The second most frequent kind of cervical cancer, adenocarcinoma, and accounts for the remaining 10 to 20% of cases [7]. The glands that produce mucus in the endocervix give rise to adenocarcinoma. While adenocarcinoma is less common than squamous cell carcinoma, it is on the rise, especially among younger women.

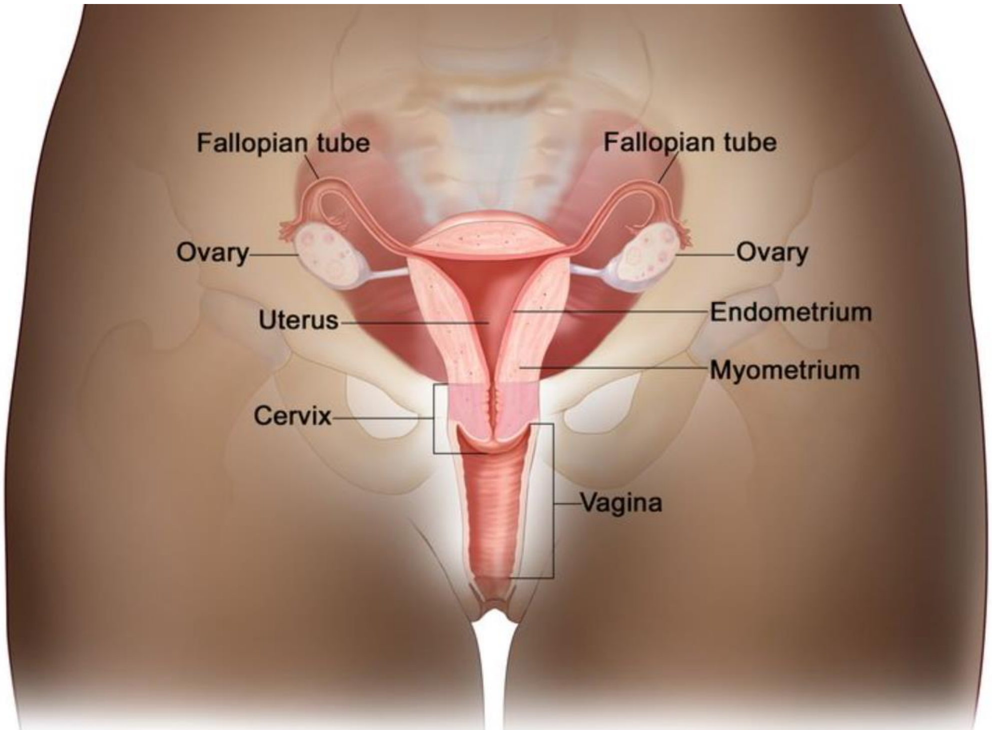
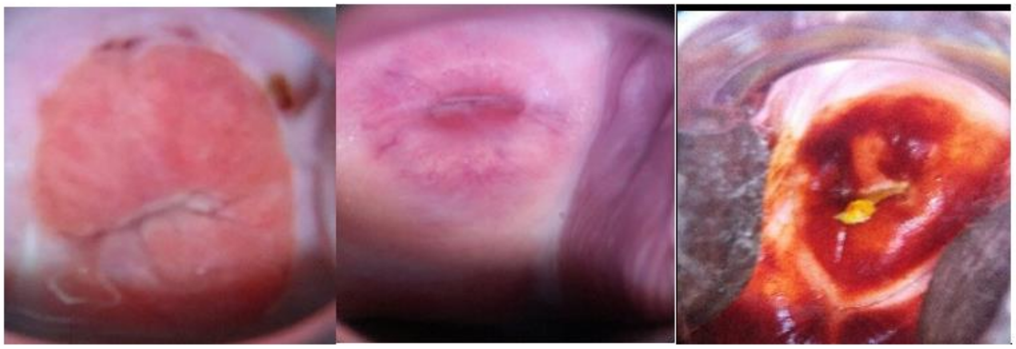


Figure 1. Cervical cancer begins in the cells of the cervix *[8]*

Cervical cell alterations that are precancerous and early malignancies of the cervix rarely elicit symptoms. As a result, regular screening with Pap and HPV tests can help detect precancerous cell changes early and prevent cervical cancer from developing. Abnormal or irregular vaginal bleeding, pain during sex, or vaginal discharge are all possible indicators of severe illness [7]. If the cancer is not treated at early it can develop to harm stage so the stage of cervical cancer is discussed under section 2.3

2.3. Cervical cancer types



type1 Type2 type 3

Figure 2. Cervical cancer stages *[9]*

Staging refers to how much cancer is present in the body and where it is located at the time of diagnosis. This is sometimes referred to as the cancer's extent. The size of the tumor, whether sections of the organ contain cancer, whether the cancer has spread from where it first started, and where the cancer has spread are all determined using information from tests and exams. The stage is used by your healthcare team to plan therapy and predict the outcome (your prognosis).

The FIGO system is the most widely used cervical cancer staging method. There are four phases of cervical cancer. The Roman numerals I, II, III, and IV are frequently used to represent stages 1 to 4. The higher the stage number, the more advanced the cancer is. Doctors may use the terms local, regional, or distant to describe the stage. The term "local" refers to cancer that has only spread to the cervix and has not spread to other areas of the body. The vaginal or pelvic regions are near to or around the cervix. Distant refers to a location on the body that is outside of the pelvic and away from the cervix. Another physician used the following terms for explaining the stages of cervical cancer type which is also in this study.

- ✓ Stages 1A, 1B, and 2A are common in early stage cervical cancer.
- ✓ Stages 2B, 3A, and 4A are common in locally advanced cervical cancer.
- ✓ Stage 4B is considered advanced cervical cancer.

Cervical cancer is divided into two types:

- ✓ Squamous cell carcinoma (SCC)
- ✓ Adenocarcinoma

Cervical cancer comes in a variety of forms. Squamous cell cancer is the most prevalent kind. The abnormality is called CIN (Cervical Intraepithelial Neoplasia)

There are three types (grades) or Stages of cancer cells. Which is known as CIN Type 1, 2, and 3 so those cervical intraepithelial neoplasia are discussed as follows.

Type-1

A cancer's stage indicates how big it is and whether it has spread. It assists your doctor in determining which treatment you require. For cervical cancer, doctors follow the International Federation of Gynecology and Obstetrics (FIGO) staging system. The stages are numbered 1 through 4. Stage 1 cancer refers to cancer that has spread to the womb's neck (cervix). It hasn't spread to surrounding organs or tissues. It is frequently subdivided into: 1st Stage 1A 1st Stage 1B

Type-2

The cancer has migrated outside the cervix and into the surrounding tissues in stage 2. Chemoradiotherapy (a mix of chemotherapy and radiotherapy), as well as surgery, are the most common therapies.

Type-3

Stage 3 indicates that the cancer has migrated from the cervix to surrounding structures or to lymph nodes in the pelvic or abdomen. Chemotherapy and radiotherapy are commonly used in treatment (chemoradiotherapy).

2.4. Digital Image processing

Image processing is the process of utilizing a digital computer to process a digital image, removing noise and any other anomalies [10]. Digital computers have become widely available and used for a variety of purposes throughout the previous four to five decades. Digital image processing is always a fascinating area since it improves pictorial information for human interpretation and image data processing for storage, transport, and representation for machine perception [11, 12].

Image processing is a technique used to improve row images received from cameras or sensors installed on aircrafts, satellites, or space crafts, or pictures taken in everyday life for a variety of applications. This field of image processing has advanced significantly in recent years and has spread to various fields of science and technology. Image processing mainly deals with image acquisition, image pre-processing, image segmentation, feature extraction, image detection and classification [13, 12]. In the following sections, we will see some image processing techniques that include image acquisition, image pre-processing and feature extraction.

2.4.1. Image Acquisition

Image processing started by image acquired using camera or scanner. Image acquisition is the process of obtaining a digital image of object in real world using camera or scanner and converted to the desired output format [12]. During image acquisition process, there exist a random change in pixel values, due to this we needs to bring the image pixel in the same format [14]. Atmosphere degradation, motion of image or camera, quality of scanner, distance between camera lens and image are some of the causes for pixel values changes in image acquisition.

2.4.2. Image Pre-processing

Image preprocessing is the major procedures to extract appropriate features from image. Regardless matter which image acquisition devices are utilized, the images with noise are always unsatisfactory. For instance, there are noises in the image, the image's region of interest is not clear, environmental elements or other objects' interference are present in the image, and so on. For different picture applications, different preprocessing methods will be chosen [15]. Picture resizing, image enhancement, segmentation, noise removal filtering, gray scale conversion, image histogram, and thresholding are all employed in this application.

2.4.3. Image Segmentation

The quality of the result is determined by picture segmentation, which is an important part of the image analysis technique. It is the technique of separating or grouping different portions of an image. The extent to which this subdivision is carried out is determined by the problem at hand [11, 16]. The image is frequently segmented during segmentation until the objects of interest are isolated from their backdrop. Picture segmentation can be accomplished in a variety of ways, ranging from simple thresholding to complex color image segmentation approaches. Normally, these components correspond to something that people can easily separate and view as independent things. Because computers lack the ability to recognize objects intelligently, numerous different methods for segmenting photos have been devised [12, 14].

The segmentation process is based on the image's numerous attributes. This could be color information, image boundaries, or a segment. Two approaches for segmenting gray level values are used in segmentation algorithms. The first is based on gray level discontinuity, which partitions a picture based on unexpected changes in gray level, while the second is based on gray level similarity, which employs thresholding and region growth. Based on the findings of the histogram analysis, the threshold value was generated, which is expected to be constant in all photographs with the same lighting conditions [17, 14].

2.4.4. Feature Extraction

The technique of extracting relevant information that can characterize an image is known as feature extraction. It can also be defined as the process of identifying an image's basic qualities or attributes. The extraction of sufficient information, which leads to the description of the kind and character of image, is one of the essential ideas in image analysis [14]. Using a combination of color and texture features to detect and classify diseases is more accurate than using each component separately.

2.5. Deep learning

Deep learning as a subfield of machine learning that uses algorithms inspired by the structure and function of the brains [18, 19]. Artificial neural networks are the models used in deep learning that are based on neural networks (ANN). Simply described, artificial neural networks are computing systems inspired by the brain's neural networks. These networks are made up of a collection of connected units known as artificial neurons or simply neurons. Each connection between these neurons can send a signal from one neuron to another, which is subsequently processed by the receiving neuron. Then it sends messages to downstream neurons attached to it, which are normally grouped in layers. Different layers may apply various transformations to their inputs and outputs. Travel from the first layer, known as the input layer, to the final layer, known as the output layer. Any layers between the input and output layers are called hidden layers. Because of numerous powerful computers (inexpensive processing units such as GPU) and a significant amount of data, neural network applications have grown quicker than ever in the recent decade. There are one or more processing levels in an ANN. Depending on the problem we want to solve the number of layers we use in the network defers. The network shallow architecture is defined as having a small number of layers, such as two or three. The network is referred regarded as deep architecture when it comprises a significant number of layers, and deep learning refers to this deep architecture of NN [19].

Deep learning entails layer-by-layer analysis of the input, with each layer extracting higher-level information about the input. Let's look at a simple picture analysis scenario. Assume that the pixels in your input image are divided into a rectangular grid. The pixels are now abstracted by the first layer. The image's edges are understood by the second layer. The following layer creates nodes from the edges. The following step would be to find branches from the nodes. The output layer will finally identify and classify the entire item. The feature extraction procedure goes from one layer's output to the next layer's input in this case. We can process a large number of features using this method, making deep learning a very powerful tool. Deep learning algorithms try to leverage the unknown structure in the input distribution to find good representations, generally at several levels, with higher level learnt features defined in terms of lower level features [18].

For decades, developing a pattern recognition or machine learning system necessitated extensive domain knowledge and meticulous hand engineering to develop a feature extractor that converted raw data (such as image pixel values) into a suitable internal representation or feature vector from which the learning system, such as a classifier, could detect or classify patterns in the input. Deep learning allows inputting the raw data (pixels in case of image data) to the learning algorithm without first extracting features or defining a feature vector. The traditional machine learning algorithm needs separate hand-tuned feature extraction before the machine learning phase. Deep learning has only one neural network phase. At the beginning of the neural network, the layers are learning to recognize the basic features of the data and that data feedforward to the other layers in the network for additional computation of the network [20, 18].

2.6. Artificial Neural Networks

The most popular and primary approach of deep learning is using “Artificial neural network” (ANN). They are brain-inspired systems that are designed to mimic how we humans learn [21]. An ANN is made up of artificial neurons, which are a collection of connected units or nodes that loosely replicate the neurons in a human brain (see Figure 3). Neurons are cells found in the central nervous system of humans. Axons and dendrites are the connecting regions between axons and dendrites, and synapses are the connecting regions between axons and dendrites. Each link allows neurons to receive data and do mathematical calculations, as well as transmit synapses represented by weights. In 1958, psychologist Frank Rosenblatt built the first Artificial Neural Network (ANN) based on this paradigm [22].

ANNs are comparable to neurons in that they are made up of numerous nodes. The nodes are firmly linked and grouped into several hidden layers. The input layer receives the data, which is then passed via one or more hidden layers in a sequential order before being predicted by the output layer.

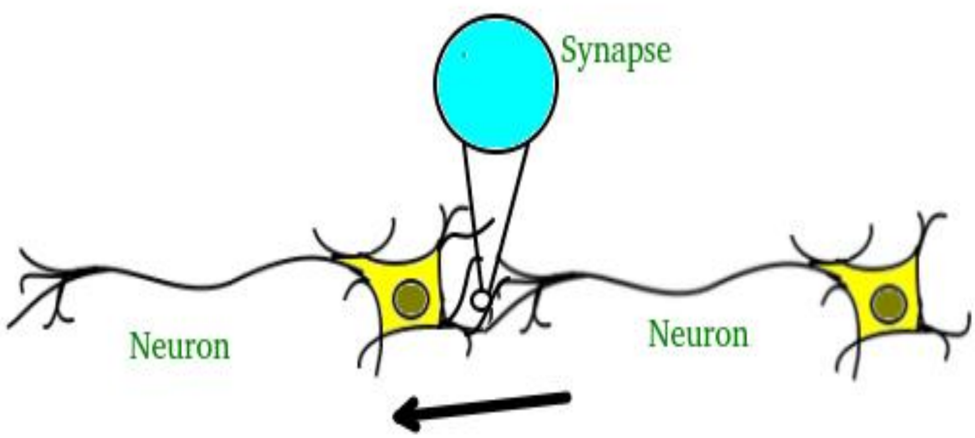


Figure 3. Example of Artificial Neurons [23]

The ANNs ability to learn non-linear complex relationship automatically from data made them a model of choice in many application areas. In addition, advances in computing power and techniques for building and training ANNs have allowed these models to achieve state-of-the-art results in image classification [24, 25, 26]. This breakthrough in image classification is attributed to a special type of ANN called CNN.

An artificial neural network computes a function of the inputs by propagating the computed values from the input neurons to the output neuron(s) and using the weights as intermediate parameters. Learning occurs by changing the weights connecting the neurons. Just as external stimuli are needed for learning in biological organisms, the external stimulus in artificial neural networks (ANN) is provided by the training data containing examples of input-output pairs of the function to be learned. For example, in thesis the training data might contain pixel representations of images (input) and their annotated labels e.g., the outputs are Type_1, Type_2, and Type_3. These training data pairs are supplied into the neural network, which makes predictions about the output labels based on the input representations. Based on how closely the projected output for a specific input matches the annotated output label in the training data, the training data offers feedback on the validity of the weights in the neural network. The errors caused by the neural network in the computation of a function might be viewed as a form of unpleasant feedback in a biological organism, causing synaptic strengths to be adjusted. In a neural network, the weights between neurons are modified in response to prediction mistakes. The purpose of modifying the weights is to modify the computed function such that future iterations' predictions are more accurate. As a result, the weights are carefully adjusted in a mathematically justified manner to minimize the computation error in that example. The function generated by the neural network is refined over time by successively altering the weights between neurons over numerous input-output pairs, resulting in more accurate predictions. As a result, if the neural network is trained with a large number of different photos of cervical cancer, it will eventually be able to correctly detect a cervical cancer in a new image. The capacity of all machine learning models to generalize their learning from observed training data to unseen samples is their major benefit. The input is modified by the weights and functions employed in the nodes as it passes through each hidden layer, eventually producing an output. When the created output is compared to the real value, the error propagates back to change the weights, learning occurs. The back propagation algorithm is what it's called. The process is repeated for a predetermined number of iterations (epochs) or until the error falls below the desired level.

2.7. Feedforward neural network

A feedforward neural network is a type of artificial neural network in which nodes' connections do not form a loop [27, 28]. As a result, it differs from its offspring, recurrent neural networks.

The feedforward neural network was the first and most basic artificial neural network to be created [19]. Figure 4 shows how information goes in this network in only one direction: forward

From the input nodes, through any hidden nodes (if any), and to the output nodes. In the network, there are no cycles or loops [28].

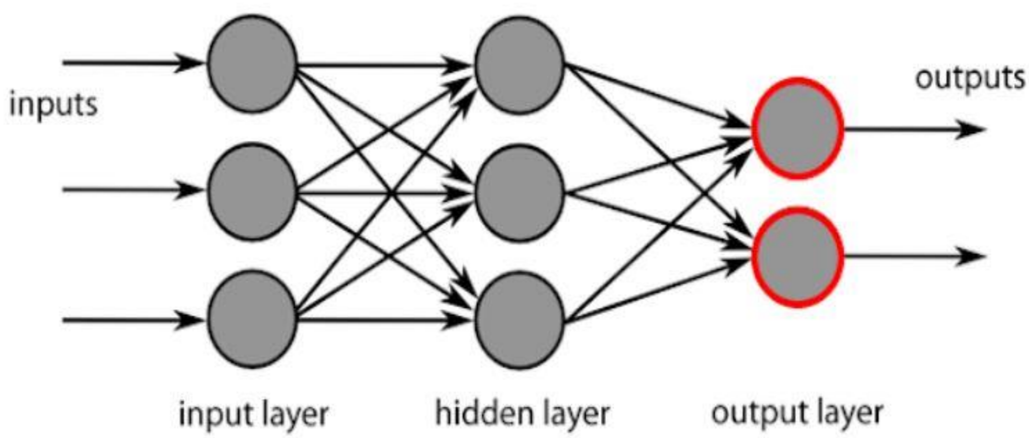


Figure 4. Feedforward neural network architecture [29]

2.8. Multi-layer perceptron (MLP)

The most basic type of ANN is the multi-layer perceptron. It has a single input layer, one or more hidden layers, and an output layer at the end [21]. A layer is made up of a group of perceptrons. One or more aspects of the input data make up the input layer. Every hidden layer contains one or more neurons that process a specific feature component before sending the processed data to the next hidden layer. The output layer procedure receives data from the last hidden layer and outputs the result. In the 1940s, Warren McCulloch and Walter Pitts described a similar neuron. As long as the threshold value is between the activated and deactivated states, any value for the activated and deactivated states can be used to generate a perceptron. To train perceptrons, a basic learning algorithm known as the delta rule can be used. It evaluates the differences between computed and sample output data and uses this information to alter the weights, resulting in a gradient descent algorithm.

Single-layer perceptron networks can only learn linearly separable patterns; Marvin Minsky and Seymour Papert demonstrated in their famous monograph Perceptron in 1969 that a single-layer perceptron network could not learn an XOR function (however, multi-layer perceptron networks can produce any possible Boolean function).

2.8.1. Activation function

It's an important aspect of a neural network's design. The activation function of the hidden layer determines how well the network model learns the training dataset. The activation function used in the output layer determines the type of predictions the model can make [30]. Each node in the neural network receives a large number of inputs and generates a single output based on the weighted sum of those inputs. This capability is achieved with a rather simple model. Frank Rosenblatt first invented the perceptron in 1957 [22]. The perceptron adds all of the inputs and their weights together, compares the result to a threshold, and then outputs a discrete value based on the comparison. Activation functions represent a key factor in ANN implementation [31].

Their major goal is to change a node's input signal to an output signal, which is why they're called transformation functions. The output of network layers without an activation function is turned into a linear function, which has several limitations due to their limited complexity and capacity to understand complicated and intrinsic correlations among dataset variables. In real-world datasets, non-linearity is common. Images, voice recordings, and videos, for example, cannot be described using this simple approach since they are made up of numerous dimensions that are not represented by linear transformations. Differentiability is another quality of activation functions that is required for a back propagation optimization technique to be successful [31]. The gradient descent optimization algorithm, which modifies the weight of neurons by computing the gradient of the loss function, uses back-propagation in the context of learning. After a training observation has propagated through the neural network, the loss function measures the difference between the network output and its predicted output.

Almost all activation functions suffer with the vanishing gradient problem, which is a well-known difficulty. As more layers with a specific activation function are added to a neural network, the gradients of the loss function approach 0 and the network training is terminated. As a result, resolving the problem becomes extremely difficult. Because back-propagation finds the network's derivatives by going layer by layer from the last to the first, the derivatives of each layer are multiplied down the network according to the chain rule statements. A small gradient (near zero) indicates that the weights and biases from the first layers will not be effectively updated while training the network, which can result in a loss of accuracy because the model is unable to recognize core elements from the input data, which occurs frequently at network entry.

Although perceptrons can theoretically be used to compute any function, they are rarely employed in practice, owing to their discontinuous nature: a little change in the weights can lead the perceptron to generate a significantly different output, which is undesirable for learning [32]. As a result, instead of using a perceptron, a sigmoid activation function is employed in this neural network implementation, which is relatively comparable to the former but eliminates its flaw. The formula for the sigmoid function is the following:

$$\boldsymbol{\sigma}(\mathbf{x}) = \frac{\mathbf{1}}{\mathbf{1+e^{x}}} \tag{2.1}$$

The function takes the following form when represented as a node in a neural network with weights, inputs, and biases:

$$\boldsymbol{\sigma}(\mathbf{x}) = \frac{\mathbf{1}}{\mathbf{1+e^{-\sum_i w_i x_i + b}}} \tag{2.2}$$

The function, by definition, takes a real-valued input and returns a value between 0 and 1. The output is closer to 1 as the input value increases, and vice versa. In this way, the sigmoid function is comparable to the perceptron in that it provides an output that is extremely close to 1 or 0.

2.8.2. Loss function

The neural network used in this thesis is designed to learn how to classify photos. To learn something, one must be given feedback on his current performance. The loss function's job is to evaluate the neural network's accuracy. If the loss is small, the neural network is correctly classifying the images; if the network is incorrectly classifying the images, the loss will be large. The output of the neural network must first be translated as class scores before the loss for a specific estimate can be calculated. The score function does this by calculating the chance that a given input is a specific class using the values from the output layer nodes. The softmax function, which is used in this thesis, is given by the following formula:

$$\mathbf{p_i} = \frac{\mathbf{e^{z_i}}}{\sum_{\mathbf{j \in group}} \mathbf{e^{z_i}}} \tag{2.3}$$

Where z is a vector of output nodes and group is a collection of node indexes in the output layer [32].

2.8.3. Gradient descent

We defined a loss function in the previous section, which essentially tells us how accurately the neural network can classify data. Now we need a way to act on this information and somehow minimize network loss during training by changing the weights of different nodes. The gradient descent algorithm is used to accomplish this.

Gradient descent works by determining the gradient of the loss function with respect to the network weights, which tell us which direction the function increases the most. To minimize the loss, a fraction of the gradient must be subtracted from the corresponding weight vector.

2.8.4. Back propagation

Back propagation is a practical realization of the gradient descent algorithm in multilayered neural networks. It calculates the gradient of the loss function with respect to all the weights in the network by iteratively applying the multivariable chain rule.

Applying the back propagation algorithm to a neural network is a two way process: we first propagate the input values through the network and calculate the errors, and then we back propagate the errors through the network backwards to adjust the connection weights in order to minimize the error. The algorithm calculates the gradient of the loss function with respect to the weights between the hidden layer and output layer nodes, and then it proceeds to calculate the gradient of the loss function with respect to the weights between the input layer and hidden layer nodes. After calculating the gradients, it subtracts them from the corresponding weight vectors to get the new weights for the connections. This process is repeated until the network produces the desired outputs.

In the neural network implemented in this thesis, the gradient of the loss function with respect to the weights between the hidden layer and output layer nodes can be computed as follows:

$$\frac{\partial \mathrm{loss}}{\partial \mathrm{w}_2} = \frac{\partial \mathrm{loss}}{\partial \mathrm{z}} \frac{\partial \mathrm{z}}{\partial \mathrm{w}_2} \tag{2.4}$$

Where loss is the cross-entropy loss described in section 2.8.2 and z is a vector that holds the values of the output layer nodes. Going further down the line, the gradient of the loss function with respect to the weights between the input layer and hidden layer nodes is calculated with the following formula:

$$\frac{\partial \mathrm{loss}}{\partial \mathrm{w}_1} = \frac{\partial \mathrm{loss}}{\partial \mathrm{z}} \frac{\partial \mathrm{z}}{\partial \mathrm{y}} \frac{\partial \mathrm{y}}{\partial \mathrm{w}_1} \tag{2.5}$$

Where y is a vector that holds the values of the hidden layer nodes.

2.9. Convolutional Neural Network (CNN)

Convolutional neural network also known as ConvNet or CNN is one of the most popular ANN [18, 33]. CNNs are a subtype of neural network that is used to process data with a known, grid-like topology, such as time series and image data [34]. Convolutional neural networks are biologically inspired networks that are used in computer vision for image classification and object detection [35]. It is a multi-layered deep learning and neural network technique. It is widely employed in image and video recognition. Its foundation is the mathematical concept of convolution. It is similar to a multi-layer perceptron in that it has a series of convolution layers and a pooling layer before the fully connected hidden neuron layer. CNN's architecture and training procedure A CNN is composed of several building blocks, which are detailed in

Section 2.9.1. Convolution layers, pooling layers (e.g., max pooling), and fully connected (FC) layers are examples. The performance of a model with specific kernels and weights is calculated with a loss function on a training dataset using forward propagation, and learnable parameters, i.e., kernels and weights, are updated according to the loss value using the gradient descent optimization algorithm described in section 2.8.3. In computer vision, images can be filtered by using convolution operation to produce different visible effects. CNN has convolutional filters that are used to detect some objects in a given image such as edges which is the same as the biological receptive field. Since the late 1980s and in 1990 CNN gives interesting results in handwritten digit classification and face recognition. The following figure (Figure 5) [36] illustrates CNN architecture.

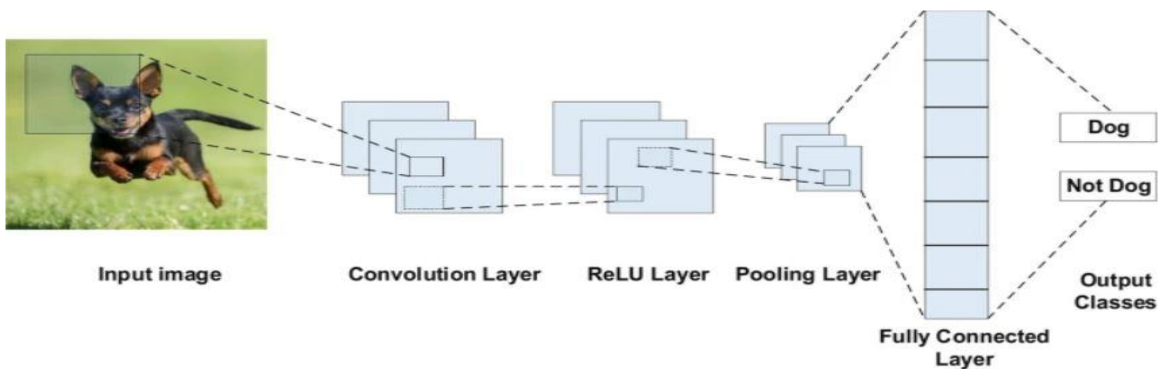


Figure 5. An Example of CNN Architecture

2.9.1. Building blocks of CNN architecture

As shown in Figure 10, the CNN architecture consists of several building blocks, including convolution layers, pooling layers, and fully connected layers. A typical architecture consists of several convolution layers, followed by a pooling layer that is repeated several times, and one or more fully connected layers. The process of transforming input data into output data using these layers is referred to as forward propagation. Although the convolution and pooling operations described in this section are for two-dimensional (2D) CNN, they also apply to three-dimensional (3D) CNN.

In this thesis, CNN is used to detect and classify Type_1, Type_2, and Type_3 cervical cancer by giving cervical cancer image as an input. For the process of detection and classification, CNN is used which is composed of various sequential layers and every layer of the algorithm transforms one volume of activation to another using different functions. The basic and commonly used layers of CNN are the convolution layer, the pooling layers, and the fully connected Layer.

I. Convolution layer

It is the fundamental building block that performs computational tasks using the convolution function. The convolution layer of a CNN architecture is a critical component that performs feature extraction. It is usually made up of a mix of linear and nonlinear operations, such as convolution and activation functions [26] Convolution is a type of linear operation used for feature extraction that involves applying a small array of numbers known as a kernel across a tensor array of numbers as input.

A feature map is obtained by calculating an element-wise product between each element of the kernel and the input tensor at each location of the tensor and summing it to obtain the output value in the corresponding position of the output tensor, as shown in Figure 6. This procedure is repeated with an arbitrary number of kernels to produce an arbitrary number of feature maps representing different properties of the input tensors; different kernels can thus be viewed as different feature extractors. The size and number of kernels are two key hyper parameters that define the convolution operation. The former is usually 3 × 3, but it can also be 5 × 5 or 7 ×7. The depth of the output feature maps is determined by the latter, which is arbitrary. The following figure (Figure 6) [37] illustrates example of convolution layer operation.

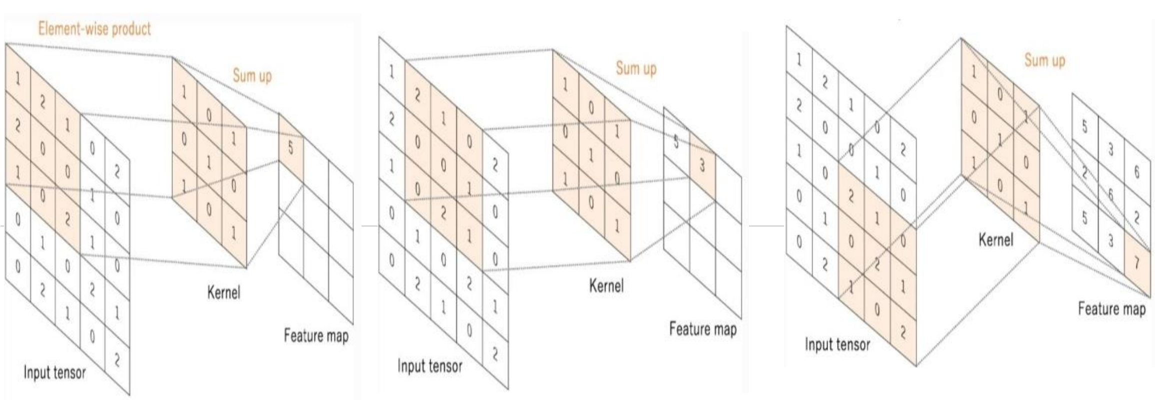


Figure 6. An example of Convolution layer operation

Figure 6 shows a convolution operation with a kernel size of three, no padding, and a stride of one. A kernel is applied across the input tensor, and an element-wise product between each element of the kernel and the input tensor is calculated and summed at each location to obtain the output value in the corresponding position of the output tensor, which is referred to as a feature map.

II. Pooling layer

It is placed next to the convolution layer and is used to reduce the size of inputs by removing unnecessary information so that computation can be performed more quickly. It implements a standard down sampling operation that reduces the in-plane dimensionality of the feature maps in order to introduce translation invariance to small shifts and distortions and reduce the number of subsequent learnable parameters. It is worth noting that none of the pooling layers have learnable parameters, whereas filter size, stride, and padding are hyper parameters in pooling operations, similar to convolution operations [26]. Figure 7 shows an example of a max pooling operation with a filter size of 2×2, no padding, and a stride of 2, which extracts 2× 2 patches from the input tensors, outputs the maximum value in each patch, and discards all other values, resulting in a factor of 2 down sampling of an input tensor's in-plane dimension. The following figure (Figure 10) [37] illustrates example of max pooling operation.

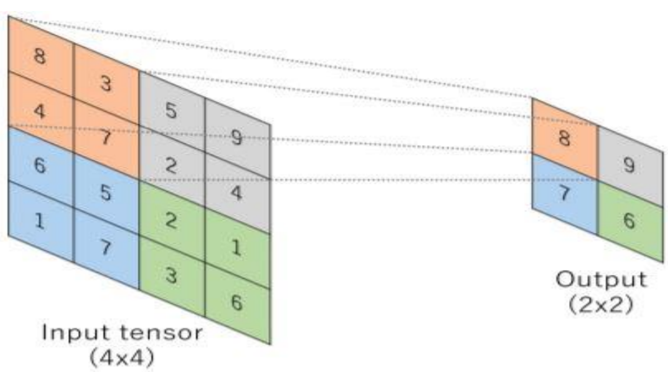


Figure 7. An example of max pooling operation

Different types of pooling methods are available for use in various pooling layers. These methods include tree pooling, gated pooling, average pooling, min pooling, max pooling, global average pooling (GAP), and global max pooling. As illustrated in Figure 11, the most well-known and widely used pooling methods are max, min, and GAP pooling. The following figure (Figure 8) [36] illustrates example of three type of pooling operations.

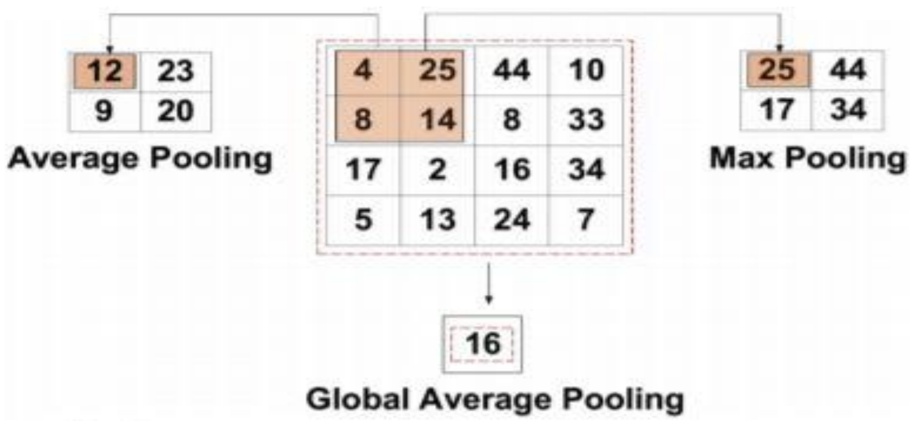


Figure 8. An Example of three type of pooling operations

III. Fully connected layer

The final pooling layer's output is flattened and fed into one of several fully connected layers that act as classifiers. CNN training is similar to fully connected neural network training that employs the back propagation algorithm. It divides input into groups and is arranged next to a series of convolution and pooling layers. The final convolution or pooling layer's output feature maps are typically flattened, i.e. transformed into a one-dimensional (1D) array of numbers (or vector), and connected to one or more fully connected layers, also known as dense layers, in which every input is connected to every output by a learnable weight. In classification tasks, a subset of fully connected layers maps the features extracted by the convolution layers and down sampled by the pooling layers to the network's final outputs, such as the probabilities for each class. Typically, the final fully connected layer has the same number of output nodes as the number of classes. Following each fully connected layer is a nonlinear function, such as ReLU.

2.9.2. Evolution of Convolutional Neural Network Models

AlexNet: The first breakthrough came in 2012 when the convolutional model which was named AlexNet significantly outperformed all other conventional methods in ImageNet Large-Scale Visual Recognition Competition (ILSVRC) 2012 that featured the Image Net dataset. The AlexNet brought down classification error rate from 26 to 15%, a significant improvement at that time. AlexNet was simple but much more efficient than LeNet [38]. The improvements of AlexNet is Large labeled image database (ImageNet), which contained around 15 million labeled images from a total of over 22,000 categories, was used and The model was trained on high-speed GTX 580 GPUs for 5 to 6 days. The other improvements are ReLU (Rectified Linear Unit) f (x) _max(x, 0) activation function was used. This activation function is several times faster than the conventional activation functions like sigmoid and tanh. The ReLU activation function does not experience the vanishing gradient problem. AlexNet consists of five convolutional layers, three pooling layers, three fully connected layers, and a 1000-way softmax classifier [18].

ZFNet: In 2013, an improved version of CNN architecture called ZFNet was introduced [39]. ZFNet reduced the filter size in the first layer from 11×11 to 7×7 and used a stride of 2 instead of 4 which resulted in more distinctive features and fewer dead features. ZFNet turned out to be the winner of ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) 2013.

VGG: introduced in 2014, Winner of the ImageNet ILSVRC-2014 competition, VGGNet was invented by Oxford's Visual Geometry Group used increased depth of the network for improving the results. The depth of the network was made 19 layers by adding more convolutional layers with 3×3 filters, along with 2×2 max-pooling layers with stride and padding of 1 in all layers. Reducing filter size and increasing the depth of the network resulted in CNN architecture that produced more accurate results. VGGNet achieved an error rate of 7.32% in ILSVRC 2014 and was the runner-up model in ILSVRC 2014 [18].

GoogLeNet: Google developed a ConvNet model called GoogLeNet in 2015. GoogLeNet proposed a novel concept referred to as inception architecture [35]. The model has 22 layers and was the winner of ILSVRC 2015 for having the error rate of 6.7%. The previous ConvNet models had convolution, and pooling layers stacked on top of each other but the GoogLeNet architecture is a little different. It uses an inception module which helps in reducing the number of parameters in the network. The inception module is actually a concatenated layer of convolutions (3 × 3 and 5 × 5 convolutions) and pooling sub-layers at different scales with their output filter banks concatenated into a single output vector making the input for the succeeding stage. These sub-layers are not stacked sequentially but the sub-layers are connected in parallel. To compensate for the additional computational complexity caused by extra convolutional operations, 1 1 convolution is used, resulting in fewer computations before expensive 3 3 and 5 5 convolutions are performed. Two convolutional layers, four max-pooling layers, nine inception layers, and a softmax layer comprise the GoogLeNet model. Because of the use of this unique inception architecture, GoogLeNet has 12 times fewer parameters than AlexNet [18].

ResNet: In 2015, Microsoft Research Asia proposed a 152-layer deep CNN architecture called ResNet [40]. ResNet introduced residual connections, which add the output of a conv-relu-conv series to the original input and then pass it through a Rectified Linear Unit (ReLU). In this way, the information is carried from the previous layer to the next layer and during backpropagation, the gradient flows easily because of the addition operations, which distributes the gradient. ResNet outperformed humans in classification, detection, and localization, winning the ILSVRC 2015 with an error rate of 3.6%, which is lower than the human error rate of 5–10% [18].

Inception-ResNet: A hybrid inception model which uses residual connections, as in ResNet, was proposed in 2017 [24]. This hybrid model, known as Inception-ResNet, significantly improved the training speed of the inception model and outperformed the pure ResNet model by a narrow margin.

Xception: is a convolutional neural network architecture based on depth-wise separable convolution layers [24]. The Xception is said to perform slightly better than the Inception V3 on ImageNet [18].

Table 1 Classification performance of VGG -16, ResNet-152, Inception V3 and Xception on ImageNet.

Model	Model Top-1	Top-5 accuracy
AlexNet	0.625	0.86
VGG-16	0.715	0.901
Inception	0.782	0.941
ResNet-152	0.870	0.963

2.10. Recurrent Neural Network (RNN)

Recurrent Neural Networks (RNN) is beneficial for addressing flaws in other artificial neural network (ANN) models. In training, the majority of the ANN didn't remember the steps from earlier cases and instead learnt to make decisions based on context. Meanwhile, RNN saves historical data and bases all of its choices on what it has learned in the past.

This method is most beneficial when it comes to image classification. In order to fix the past, we may need to reach into the future. Bidirectional RNN is useful in this scenario for learning from the past and predicting the future. We have handwritten examples in several inputs, for example. If one of the inputs is confusing, we must review the other inputs again to identify the proper context, which is based on a decision made in the past.

RNN is one of the first deep learning architecture which gives a road map to develop other deep learning algorithms. It is commonly used in speech recognition and natural language processing [41]. RNN is designed to recognize the sequential characteristics (remembers previous entries) of the data. When we analyze time serious data, the network has memory (hidden state) to store previously analyzed data. To perform the present task RNN needs to look at the present information (short term dependency) and this is the main drawback. RNN differs from a neural network is that RNN takes a sequence of data defined over time [41].

LSTM is a special type of RNN which is explicitly designed to overcome the problem of long-term dependencies by making the model remember values over arbitrarily time interval. The main problems of RNN are vanishing gradient and exploding gradients. The gradient is the change of weight with regard to the change in error. It is well suited to process and predicts time series given time lags of unspecified duration. For example, RNN forgets the model if we want to predict a sequence of one thousand intervals instead of ten, but LSTM remembers such kind of activities. The main reason that LSTM can remember its input in a long period of time is that it has a memory that is like memory on a computer which allows the LSTM to read, write and delete information [42]. It is mostly applied to natural language text compression, speech recognition, handwritten recognition, gesture recognition, and image captioning.

Recurrent neural networks have improved long short-term memory (RNNs). In order to solve the vanishing and exploding gradient problem, LSTM suggests memory blocks instead of traditional RNN units. The key difference between it and RNNs is that it adds a cell state to save long-term states. An LSTM network can remember and connect data from the past with data from the present. The LSTM is made up of three gates: an input gate, a "forget" gate, and an output gate, with x_t denoting the current input, C_t and C_{t-1} denoting the new and previous cell states, and h_t and h_{t-1} denoting the current and prior outputs, respectively. Figure 9 depicts the internal structure of the LSTM. The following diagrams depict the LSTM input gate's principle.

$$i_t = \sigma(W_i . [h_{t-1}, x_t] + b_i) \tag{2.6}$$

$$\widetilde{C}_t = \tanh(W_i . [h_{t-1}, x_t] + b_i) \tag{2.7}$$

$$C_t = f_t C_{t-1} + i_t \widetilde{C}_t \tag{2.8}$$

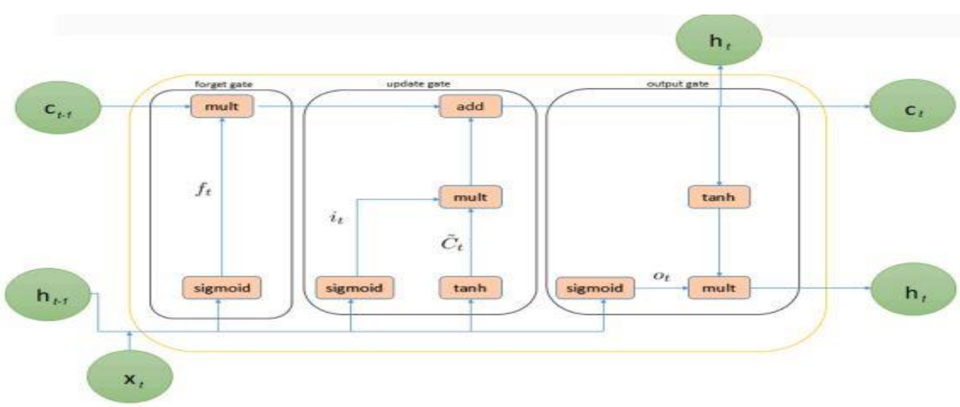


Figure 9. The internal structure of long short-term memory.

2.11. Related works

This quantitative descriptive quantitative study stated that cervical cancer is one of the disease with the highest mortality rate. So to reduce this mortality rate the study can be an alternative to help medical authorities in diagnosing cervical cancer with the help of computer aided diagnosis system. To achieve this goal they conducts two process called preprocessing and classification. The colposcopy image obtained from Guneva foundation was augmented to 500 data split into 400 and 100 for training and testing respectively. The preprocessing stage such as grayscale, histogram equalization and media filter are applied on those data and delivered as an input for classifier. The classification process carried out by using DBN (Deep Belief Network) method. DBN method is a method generally requires large amount of data labeled balanced. The result for identification of cervical cancer was depend on the accuracy achieved. The highest accuracy achieve highest identification. So the deep belief network achieves 84 percent accuracy on 200 hidden nodes [43][33].

This study discusses that cervical cancer is a cancer being one of pressing issues that need a focus. Even if there are many diagnosis techniques for a cancer such as screening the risk factors, different machine learning approaches , the paper used an approach using ensemble methods with SVM, Decision Tree, Multilayer Perceptron and K-nearest neighbors were a machine learning algorithms used with WEKA data mining tool for diagnosis purpose. 858 patients data set with 36 attributes were used for the classification was collected from the UCI machine learning repository. This paper stated that Ensemble method are hybrid algorithms that combine different machine learning algorithm into one predictive model. Finally SMO had a better performance than the others. Also the paper recommends that the accuracy can further be improved with nature inspired optimization algorithms [44].

This paper determine that early detection and assessment can increase the chances of remission due to the decelerated growth of cervical cancer. The study employs data mining technique on a patient information to predict the availability of cervical cancer. In this study data mining steps such as clustering, classification and regression are processed to determine factors which can affect a patients chances of developing cancer through Boruta Analysis. Past medical record and habits of total 858 patients with 30 attributes were collected at hospital. So finally the study conclude that the people have an alternative solutions to test for cervical cancer because of freely sourced and easy-for-all data mining. And also that the number of cigarettes smoked as well as various STDs are some of the major factors in testing for cervical cancer [45].

In this paper the issue of cervical cancer is one of the most lethal disease is raised. Early treatment of cancerous patient is proofed effective to decrease the lethal rate of the disease. So to solve this problem the study creates a deep learning model based on capsule. The dataset used for the study is open source data from kaggle.com. A type of data was colposcopy images with the number of 8141 in .JPG format. It was augmented with 20% to 25% and split into 90% as the training, 5% as the validation test and 5% as the test dataset. Finally as a result the proposed model can produce up to 94.98% accuracy. This paper recommends that the accuracy of the proposed model can be improved by developing a more in-depth architecture. As well as making better hyper-parameters and using a larger quantity of datasets [46].

In this study accurate and timely cancer detection can save lives has been discussed. So this can be devised through the accurate segmentation and classification of Pap smear cell images. The aim of this paper was to develop a better system for the automatic detection of cancer cells using a deep learning approach on Pap smear images. To achieve the aim 917 Pap smear ready-made dataset from Herlev University hospital was used to evaluate the proposed model. In implementation part the dataset is comprised in two steps to classify as normal and abnormal. Mask R-CNN segmentation in first stage and ResNet 10 was employed in second stage for for classifying. Finally in the segmentation phase mask R-CNN outperforms the previous segmentation method in precision (0.92+- 0.06), recall (0.91+-0.05), ZSI (0.91+-0.04). In the classification phase VGG- like Net yields sensitivity score of more than 96%. Lastly the paper recommends that to use of a deeper network to improve the performance result [47].

Table 2. Summary of Related Work

Author(s) and Year	Techniques used	Application(Focus area)	Dataset	Results

[33], 2020	Preprocessing(greyscale, histogram equalization and media filter)	Detection and Classification of cervical cancer	500 Ready-made dataset	84% Classification accuracy
[50],2019	Ensemble method with bagging and boosting	Prediction	858 ready-made historical	98.12% Accuracy
[51], 2019	Data mining technique(Boruta analysis)	Detection and preventation of cervical cancer	858 ready-made historical	
[52],2019	Capsule Networks	Classification	8141 images from kaggle.com	94.98% Accuracy
[53], 2019	Mask R-CNN- for segmentation VGG-like NET for classification	Segmentation and classification of cervical cancer infected class	917 pap smear image	96% classification Accuracy

CHAPTER THREE

3. RESEARCH METHODOLOGIES

3.1. Introduction

A detection and classification of cancer in to its appropriate classes becomes a wide area of research. The detection and classification of cancer study passes through a series of steps/procedures that would be applied to differentiated cancer disease. This chapter focuses on the description of methodologies that are used in order to accomplish this thesis including methods to implement the model, data collection, data preparation, software, and hardware configuration of the system used and evaluation techniques which are used to evaluate the model. In this thesis experimental research approach is used. Different experiments are carried out by using different dataset ratio. In addition to this there are a lot of experiments conducted by using different learning rate and hyper parameters.

3.2. Research Flow

In this thesis experimental research method is followed. In order to achieve the objective of this thesis, the following process flow of Figure 10 is followed. As we can see in the following Research flow, this thesis is conducted with three main phases. The first phase reviewing different kinds of literature to identifying and understanding the domain of the problem. Then objectives of the thesis are formulated including the general and specific objectives. The second phase is about data preparation and design of the thesis. During data preparation a Colposcopy image data is collected from Betazata hospital and add some downloaded image data from publicly available of kaggle site, finally splitted in to training, validation, and testing. After data preparation, design of the model is performed. The third phase is about implementation of the thesis, in this phase the designed model is implemented with appropriate tools and methods. The designed model is trained and tested with the appropriate data. During the training of the model the performance of the model is evaluated. After getting the optimal model during evaluation, the model is tested with test data. Finally, the model is compared with other different models.

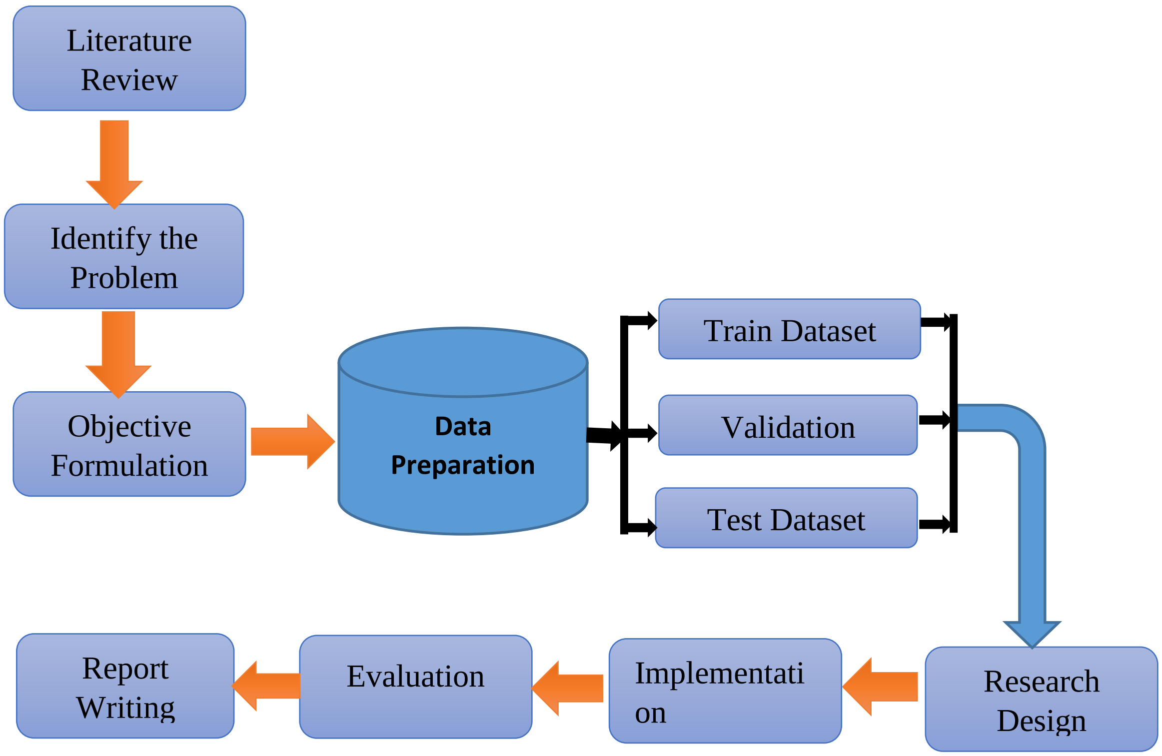


Figure 10. Research flow

3.1.1. Data Set Preparation

The most important step in using a neural network or deep learning algorithm in research is to collect the data that will be used to train the model. Imaging data for cervical cancer is the main input to the model in this thesis. A colposcopy image data were collected from Bethazeta Hospital in Addis Ababa Ethiopia and some of data was added from public dataset.

Collecting such colposcopy image is a more cumbersome job due to less data handling culture in Ethiopia. And also there is lack of the equipment and expert in using this effective method to detect the cancer. But after all, we prepare 2085 image dataset for detecting and classify the cervical cancer for three classes (Type_1, Type_2 and Type_3).

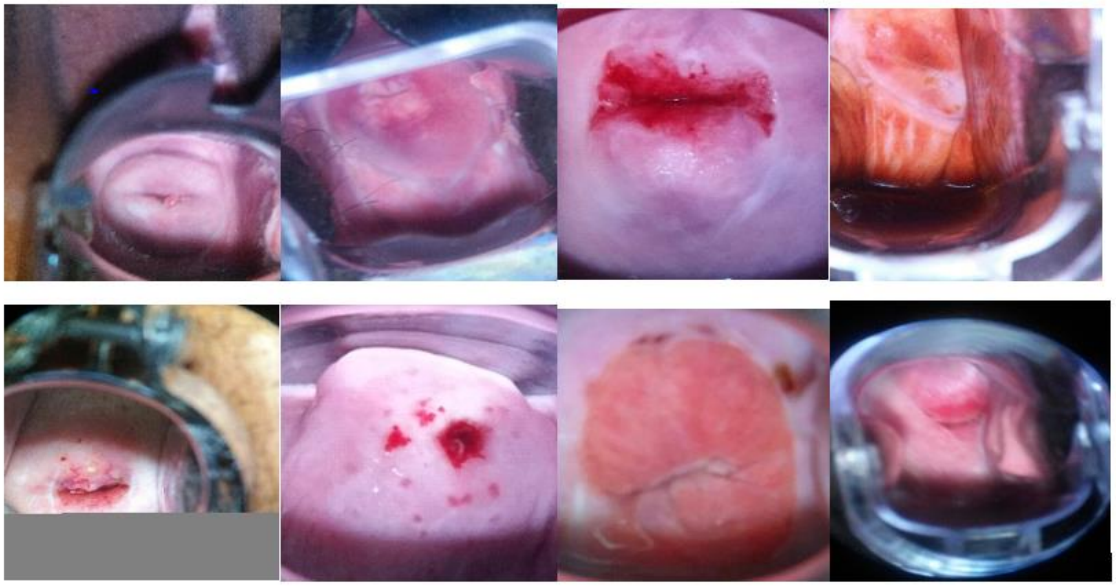


Figure 11. Example of Colposcopy Images.

3.3. Data Partitioning

First and foremost, the dataset is divided into two parts: training and testing. The training split is used to train the model, while the test split is used to test the model after it has been trained. The validation split is used to evaluate the performance of the model built during training and to select the best performing model. The original images collected from the hospital a Colposcopy image data and add some image data from publicly available of kaggle site totaled 2085. In this thesis, experiments are carried out with two different ratios: 7:3 and 8:2. Finally, the best ratio of the size of the training image to the size of the testing image was 8:2, which means that 80 percent of the dataset is for training and 20 percent is for testing. Twenty percent of the images from the training split are used for validation. As a result, the training dataset has 1668 images and the validation dataset has 417.

3.4. Image Augmentation

To perform well, deep networks require a large amount of training data. Image augmentation is typically required to improve deep network performance in order to build a powerful image classifier with little training data. Another common preprocessing technique is to add perturbed versions of existing images to the existing data set. Scaling, rotations, and other affine transformations are commonly used. This is done in order to subject the neural network to a wide range of variations. This reduces the likelihood that the neural network will recognize undesirable characteristics in the data set. Data augmentation is a process of increasing the number of training data points in a dataset by generating more data from the existing training sample [48]. It is important to increase the number of data points even if there is a large data set. It helps the network to learn more complex features from the data and prevents the problem of overfitting [48]. In this thesis, data augmentation techniques were applied to the original images in order to obtain additional images for our data set. We can augment the data during training. In this thesis, data augmentation is performed during network training using Keras libraries. The original image generates each image that was sent into the network during training.

3.5. Software Tools

The available software tools and their libraries are investigated in order to select the best tool for implementing the CNN algorithm for enset image classification. During the investigation, we discovered that there are tools that are general for both deep learning and machine learning algorithms, as well as tools that are specific to one of them. Before selecting the tools, we considered some criteria that will help us select the appropriate software tools and libraries. The main criteria are the programming language that will be used to implement the algorithm. The other criteria are to select tools with sufficient learning materials, such as free video tutorials, existing experience, and to use the tools in machines with limited resources (like CPU only). We used Python as a programming language with the TensorFlow and Keras libraries on an Anaconda environment to implement the CNN algorithm. These tools meet all of the criteria for consideration, and they are written in Python, which we are familiar with.

Anaconda it is a free and open-source distribution of the Python and R programming languages for data science and machine learning applications, with the goal of simplifying package management and deployment. It comes with a number of IDEs for writing code, including Jupyter Notebook and Spyder. Jupyter notebook was used to implement the coding. It is simple to use and runs in a web browser.

TensorFlow is the most well-known and quick deep learning library at the moment, and it was created by Google as a free and open-source library. [35]. It can be used in the cloud as a service, on mobile devices like iOS and Android, and on any desktop computer running Windows, macOS, or Linux. TensorFlow's architecture is effective for data preprocessing, model construction, model training, and model estimation. Tensors (an n-dimensional array) are used in every computation in TensorFlow to represent various types of data. For the graphical representation of the series of computations during training, TensorFlow also uses a graph framework. It has two CPU and GPU distributions.

Keras It is a free and open-source distribution of the Python and R programming languages for data science and machine learning related applications, with the goal of simplifying package management and deployment. It includes a variety of IDEs for writing code, including Jupyter Notebook and Spyder. To implement the coding, we used Jupyter notebook. It is simple to use and runs in a web browser [35].

3.6. Hardware Tools

To implement the deep learning algorithm type with the selected software tools a very slow machine with CPU Intel(R) Core (TM) i4-5200 CPU @ 2.20GHz processor, memory 8 GB and 1 Tera byte was used, and no GPU which is the most important hardware in deep learning for computer vision research.

3.7. Evaluation Methods

One of the performance evaluation method is confusion matrix. Confusion matrix is a very popular measure used while solving classification problems. It can be applied to binary classification as well as for multiclass classification problems. Confusion matrices represent counts from predicted and actual values [49]. It shows True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). True Positive indicates the number of positive examples classified accurately. Similarly, True Negative shows the number of negative examples classified accurately. Further, False Positive value shows the number of actual negative examples classified as positive and False Negative value provides the number of actual positive examples classified as negative. One of the most commonly used metrics while performing classification is accuracy. The accuracy of a model (through a confusion matrix) shows an overall performance of the system proposed by the study [49].

Accuracy can be misleading if used with imbalanced datasets, and therefore there are other metrics based on confusion matrix which can be useful for evaluating performance. There are several evaluation metrics for classifiers such as precision, recall, and the F-1 score [50]. For Precision, we think about predictions (How often is it correct?) as our classes, for recall think about truth as our base and F-Score is a weighted average of the true positive rate (recall) and precision.

CHAPTER FOUR

4. MODEL DESIGN AND EXPERIMENTS

4.1. Overview

This chapter focuses on the architecture of the selected model and its experimental setups. The design and descriptions of layers found in the selected model discussed along with how the model extract features from the input image to detect and classify the target object.

4.2. The Model Selection

CNN is a deep learning algorithm that was chosen based on research done in computer vision, specifically image classification. CNNs are a fascinating method for adaptive image processing. The algorithm is used for feature extraction, classification, training, testing, and determining the model's accuracy. CNNs process raw data without the need for additional pre-processing or feature extraction. Furthermore, the stages of feature extraction and classification occur naturally within a single framework.

The CNN algorithm is more robust and automated than traditional machine learning algorithms for the detection and classification of cervical cancer [35]. There is a need to develop different algorithms for different problems in classical machine learning algorithms, so it uses more handcrafted algorithms, but in CNN, we developed an algorithm for the detection and classification of cervical cancer. CNN will be used in this thesis for several reasons, including:

- A lot of previously conducted researches has shown that CNN is better than other classification algorithm and it is state of the art for computer vision applications.
- CNNs are designed to mimic humans' understanding of vision, so they outperform other deep learning models for image-related tasks.
- Before classification and prediction, the majority of traditional machine learning techniques demand the explicit extraction of the image features that are used for analysis.
- The majority of real-world images are tensors (3 dimensional), not vectors, so most neural network algorithms require that the actual image (input) be flattened

to a 1D vector, which is very challenging and computationally expensive. CNN, however, accepts 3-dimensional images.

- CNN can capture temporal and spatial dependencies with the help of relevant

4.3. Overview of cervical cancer detection and classification

The system design starts with the preparation of a dataset used for training and testing the model. The Cervical Cancer diseases on a colposcopy collected at the Hospital was pre-processed. Collected images filtered and categorized into their correspondence stage type. Cervical Cancer stage such as Type_1, Type_2, Type_3 and Normal had been identified. However, due to lack of data on cancer like stage 1 Normal was not included. Data resizing to 64*64 to reduce the training time of a model in detection and classification. The labeled image split into training and testing data. Then the data augmented for reducing overfitting of the model. Then developing the model from the pre-trained model by customizing a prepared cervical cancer dataset. The model evaluated with a test data set and its inference exported for delivering a predictive model and finally the predictive model.

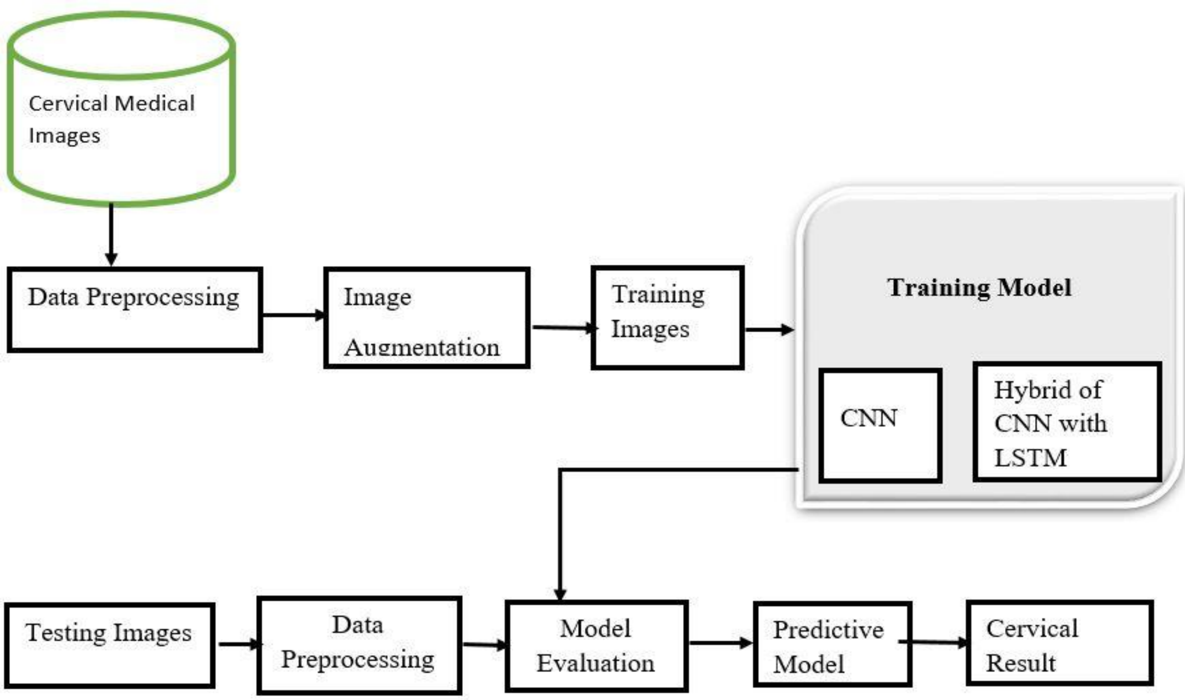


Figure 12. Block diagram of the detection and classification of cervical cancer

4.4. Training Components of the Proposed Model

The selection of CNN architecture is a very difficult part because most of the architectures are deployed in large scale applications such as ILSVRC which contains millions of parameters with thousands of classes and needs high computational power. In this thesis, the architecture is deployed in limited hardware resources and which is designed for only two classes. In order to find an appropriate model, a CNN model is designed which will work pretty in a small number of images with very low computational resources like CPU and GPU. The proposed model has 5 convolution layers, 3 fully connected layers, ReLU in the hidden layers is included as activation function to add nonlinearity during the training of the network, and dropout is included after the first two fully connected layers to prevent the problem of overfitting. Due to the reduction of trained classes, hardware resources that we have, and a number of images, we scaled down the number of neurons, parameters, and filters of pre-trained CNN models.

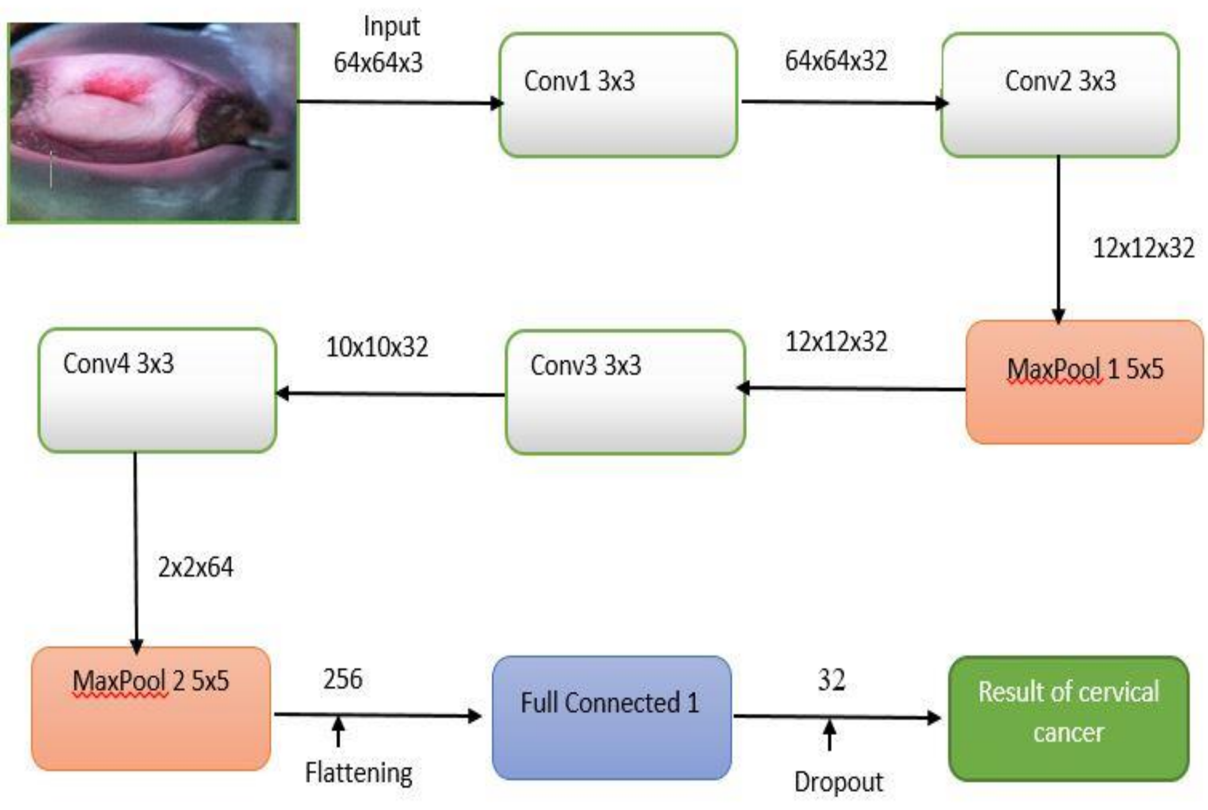


Figure 13. Proposed model for cervical cancer

4.5. Experimental Setup

In our experiment we have used three different model with similar amount of dataset to train and validate model and also three learning rate with image augmentation hyper parameters.

4.5.1. Training Components of the Proposed CNN Model

The architecture of a CNN varies depending on the problem at hand. By training numerous models with different architectures and network characteristics like learning rate, kernel size, and filter size, we were able to develop the model by trial and error. After each convolutional layer, we applied the ReLU activation function, which has been shown to result in faster training [25]. The models are trained for 50 epochs using the Adam optimizer with a batch size of 32. The architecture of this thesis is built for four classes and is deployed on constrained hardware resources. To find an appropriate model, a CNN model is created that works well in a small number of images and requires very little computational resources such as CPU and GPU. The proposed model includes four convolution layers, two fully connected layers, and a ReLU in the hidden layers as an activation function to add nonlinearity during network training. We reduced the number of neurons, parameters, and filters of pre-trained CNN models due to a reduction in trained classes, hardware resources, and the number of images.

Table 3. CNN model

Layer		**Filter**	**Features(Depth)**	**Stride**	**No. of Param.**	**Output size**
Input	Image	-	-	-		64, 64, 3
1	Conv2D + ReLU	3× 3	32	2	896	64, 64, 32
2	Conv2D + ReLU	3 × 3	32	1	9248	62, 62, 32
	MaxPool2D	5 × 5	-	1	0	12, 12, 32
	Dropout	-	-	-	0	32
3	Conv2D + ReLU	3× 3	32	2	9248	12, 12, 32
4	Conv2D + ReLU	3 × 3	32	1	9248	10, 10, 32
	MaxPool2D	5 × 5	-	1	0	2, 2, 64
	Dropout	-	-	-	0	
	Flatten	-	-	-	0	128

5	FC + ReLu	-	32	-	33024	256
	Dropout	-	-	-	0	
Output	FC + Softmax	-	3	-	771	3
Total Number of Parameters					**62,435**	

The convolutional neural network consists of 4 sets of convolution and 2 pooling. After the sets of steps are completed, the resulting values are given to the hidden layer and outputs provide. As shown Table 4 the initial spatial size of the input volume is 64 × 64 × 3 and it gets changed after some convolution operations, the initial size of the filter F and Stride S are 3 × 3 × 3 and 2 respectively and these sizes are changed after some convolution and pooling operations, and there is same padding P in our network and the value of P is always same throughout the model.

During the process of resizing images in our dataset, we considered this parameter setting to be valid. If this arrangement is not followed, libraries used to implement the CNN model will throw an exception, zero pad the rest of the area, or crop the image to fit.

Input layer: the CNN model accepts RGB images of 64 × 64× 3 pixels in size, with three different classes such as Type 1, Type 2, and Type 3. This layer only sends the input to the first convolution layer and does no processing. As a result, there are no learnable features in this layer, and the number of parameters is zero.

Convolutional layer: in the CNN model there are 4 convolutional layers. The first convolutional layer of the model filters the 64 × 64× 3 input image by using 32 kernels with a size of 3 × 3 × 3 pixels. This layer has a depth of K = 32, the output volume of this layer is 64 × 64 × 32. The product of the output volume gives a total number of neurons in the layer (first conv layer) which is 62,435.

The output of the first convolutional layer is fed into the second convolutional layer, which filters it using 32 kernels of size 3 × 3 × 32 and a 5x5pooling layer. The third and fourth convolutional layers are connected and filter it using 32 kernels of size 3x3x32, and the fourth layer has a pool size of 5x5. The CNN model's convolutional layers all use ReLU non linearity as an activation function. ReLU was chosen because it is faster than other nonlinearities like tanh for training deep CNNs with gradient descent.

Pooling layer: the proposed model has two max-pooling layers between the second and fourth convolutional layers. The first max-pooling layer uses a 5 5 filter to reduce the output of the second convolutional layer. The second max-pooling layer takes the output of the fourth convolutional layer as an input and pools it using 5X 5 filters.

Fully Connected (FC) layer: there are two fully connected layers in this CNN model, including the output layer. The first fully connected layer has 256 neurons, while the second and final fully connected layer, which is the model's output layer, has four neurons. After converting the 3D volume of data into a vector value, the first FC layer accepts the output of the four convolutional layers (Flattening). This layer computes the class score and the number of neurons in the layer that was predefined during model development. It's the same as regular NN, and as the name suggests, each neuron in this layer is linked to every number in the previous layer.

Output layer: the output layer is the last (the second FC layer) of the model and it has 4 neuron with a softmax activation function. Because the model is designed to classify 3 classes called Categorical classification.

As shown in the table above, the CNN model has 62,435 parameters, which is extremely small when compared to other deep learning architectures like AlexNet, which has 60 million parameters, VGG, which has 138 million parameters, and GoogLeNet, which has 4 million parameters. Deep learning models are thought to have a massive number of parameters; thus, they require massive computational power to train those models from scratch, as well as a massive amount of data. However, the proposed model is trained with minimal resources and data and performs admirably.

4.5.2. Training Components of the Proposed Hybrid of CNN and LSTM Model

A CNN-LSTM network hybrid is used to automatically classify and detect cervical cancer images. CNN is used to extract features in this thesis, and LSTM is used to detect and classify cervical cancer based on those features. The LSTM network has internal memory that allows it to learn from long-term states of imperative experiences. In fully connected networks, layers are fully connected, and nodes between layers are connectionless and process only one input. In the case of LSTM, nodes from a directed graph are connected along a temporal sequence, which is treated as an input with a specific order. As a result, the 2-D CNN and LSTM layout feature combination improves classification significantly.

Table 4. Hybrid of CNN and LSTM model

Layer		Filter	Features(Depth)	Stride	No. of Param.	Output size
Input	Image	-	-	-		64, 64, 3
1	Conv2D + ReLU	3× 3	32	2	896	64, 64, 3 2
2	Conv2D + ReLU	3 × 3	32	1	9248	62, 62, 3 2
	MaxPool2D	5 × 5	-	1	0	12, 12, 3 2
	Dropout	-	-	-	0	32
3	Conv2D + ReLU	3× 3	32	2	9248	12, 12, 3 2
4	Conv2D + ReLU	3 × 3	32	1	9248	10, 10, 3 2
	MaxPool2D	5 × 5	-	1	0	2, 2, 64
	Dropout	-	-	-	0	
5	lambda_1 (Lambda)	-	64	-	0	2,64
6	lstm_1 (LSTM)	-	-	-	1181696	512
7	FC + ReLu	-	32	-	131328	256
	Dropout	-	-	-	0	
Output	FC + Softmax	-	3	-	771	3
Total Number of Parameters					**1,342,435**	

The hybrid of convolutional neural network and long short term memory consists of 4 sets of convolution and 2 pooling. After the sets of steps are completed, the resulting values are given to the LSTM layer, hidden layer and outputs provide. As shown Table 5 and Figure 18 shown the initial spatial size of the input volume is 64 × 64 × 3 and it gets changed after some convolution operations, the initial size of the filter F and Stride S are $3 \times 3 \times 3$ and 2 respectively and these sizes are changed after some convolution and pooling operations, and there is same padding P in our network and the value of P is always same throughout the model. We have considered this setting of parameters to be valid during the process of resizing images contained in our dataset. LSTM reshape the input size inside of layer which is accepted form the fourth convolution and second pooling layer.

4.5.3. Augmentation Parameters

On Methodologies explained the way of preprocessing technique to scaling, rotations and transformation on existing images. We have using different augmentation parameters that are described in Table 5 below the techniques help to generate additional number of image from the original or existing training image to protect overfitting. In this section, we will see some basic geometric transformation applied to images to increase the training dataset. An essential concept of when applying these transformation is that the transformations applied to labeled data do not change the semantic meaning of the label. In this thesis, taking a case in computer vision, the scaled, rotated, flipped, translated image of a cervical cancer would be still be an image of a cervical cancer, and thus it is possible to apply these transformations to produce additional training data while preserving the semantic meaning of the label.

- **Flipping:** Images are added by flipping the original image vertically or horizontally. Depending on the dataset, this may not be a label preserving transformation.
- **Rotation:** Rotation augmentations are performed by rotating the image right or left on an axis ranging from 10 to 3590. The rotation degree parameter and problem type have a large impact on the safety of rotation augmentations. In our thesis, we used a rotation range of 25.
- **Scale:** The image can be scaled up or down. The final image size will be larger than the original image size as scale outward. Most image frameworks extract a section of the new image that is the same size as the original image.
- **Translation:** just involves moving the image along the horizontal or vertical direction (or both). This can be very useful to avoid positional bias in the data.

For example, if all the images in a dataset are centered, this would require the model to be tested on perfectly centered images as well. Performing random translation on the training dataset improves generalization ability of the model without imposing this constraint.

- **Adding noise:** Over-fitting usually happens when your neural network tries to learn high frequency features (patterns that occur a lot) that may not be useful. Adding just the right amount of noise can enhance the learning capability.

Table 5. Augmentation techniques used

Augmentation parameter	Augmentation Factor
Rotation Range	25
Width Shift Range	0.1
Height Shift Range	0.1
Shear Range	0.2
Zoom Range	0.2
Horizontal flip	True

4.5.4. Hyper parameter Settings

Hyper parameters are configurations that are defined before the training process begins and are external to the deep learning algorithm. There is no universal rule for selecting the optimum hyper parameters for a particular situation [19]. As a result, many experiments are carried out in order to select the hyper parameters. Hyper parameters for the model are described in the next sections.

- **Optimization algorithms:** To minimize the error rate, the proposed model is trained using the gradient descent optimization algorithm, and the weights are updated using the backpropagation of error algorithm. Gradient descent is by far the most widely used and popular optimization algorithm in deep learning research [18]. At the same time, every cutting-edge deep learning library includes Keras implementations of gradient descent optimization algorithms. It updates the model's weight and tunes parameters to minimize the loss function. The Adaptive Moment Estimation (Adam) optimizer is used to optimize the gradient descent [48]. Adam computes adaptive learning rate for each parameter and scales learning rate using squared gradients, as well as the gradient's moving average.

- **Learning rate**: Because the proposed model was trained using backpropagation, the learning rate is used during weight update. It determines how much weight to update during backpropagation [19]. Choosing the appropriate learning rate during our experiment was the most difficult part. In our experiment, we discovered that a learning rate with a value of too small takes longer to train than a value of larger. When we give a lower value, however, the model outperforms a model with a higher learning rate. The experiment was carried out with learning rates of le-3, 0.001, and 0.1. For all of the experiments, the optimal learning rate is then Le-0.001.

- **Loss function:** The loss function chosen is directly related to the activation functions used in the model's output layer and the type of problem we are attempting to solve (whether regression or classification). Softmax is used as an activation function in the final fully connected layer in the proposed model. We are dealing with a classification problem, specifically a categorical classification problem. For our model, we used Categorical Cross-Entropy (CCE) loss as a loss function. Although other loss functions such as Binary Cross-Entropy (BCE) and Mean Squared Error (MSE) exist, categorical cross-entropy is the preferred loss function for multiclass classification [48, 35]. It works well for models with multiclass output, as it measures the distance between the actual and desired output. CCE loss was used in the experiment.

- **Activation function**: We have several activation functions. SoftMax activation functions are used in our experiments for multiclass classification. Because it is the best choice for a multiclass classification problem, the Softmax activation function is used in the model's output layer [35, 48].

- **Number of epochs**is the number of iterations that the entire dataset undergoes in the model or network, both forward and backward. In our study, the model was trained using a variety of epochs ranging from 10 to 50. We have observed that the model exhibits a significant gap between the training error and validation error when we use an epoch that is either too small or too large. The model achieves its optimal state at epoch 50 after numerous experiments.

- **Batch size**: is the maximum amount of input data we send to the network simultaneously. We must split the input into several smaller batches because it would be difficult to provide the computer with all the data in a single epoch. Minimizing the machine's computational time is preferred when training models. In our experiment, a batch size of 32 is used during model training.

Table 6. Summary of hyper parameters used during model training

Parameter	Epoch	Batch	Activation	Loss	Optimization	Learning
Value	50	32	Softmax	CCE	Adam	Le-0.001

4.6. Experimental Result

In this section, an attempt is made to construct model for cervical cancer detection and classification by using the deep learning algorithm like CNN and hybrid of CNN with LSTM and comparing both of them. In this section detailed implementation procedures, experimentation, analysis and the result are presented below. The results of the experiments are shown in tables.

The first scenarios are based on deep learning type called CNN and hybrid of CNN with LSTM model to classify the disease and comparing three of them in different hyper parameters. Like most of the deep learning classification algorithms, our experiments have two main phases. The first one is the training phase and the second one is the testing phase. In the training phase, data is repeatedly presented to the classifier, while weights are updated to obtain the desired response. In the testing phase, the trained algorithm is applied to data that has never seen (test data) by the classifier to test the performance of the classification algorithm. In the following, we will see the experimental results in detail.

4.6.1. Detection and Classification of Cervical Cancer using the CNN and Hybrid of CNN and LSTM Model

The model is trained with a total of 2085 original images before data augmentation during the training is applied in the dataset. A lot of experiments are conducted by the proposed model by changing the ratio of training and testing dataset, by using different learning rates, and finally by softmax activation functions.

4.6.1.1. Experimental 1: Changing Training and Testing Dataset Ratio for CNN and Hybrid of CNN and LSTM Model

The outcome of an experiment with the CNN and Hybrid of CNN and LSTM model using image data split for training and testing. Choose ratios for training and testing from the given data set for our experiment. The ratio 8:2 used for training and testing, whereas the ratio 7:3 used for training and testing. For the testing data set, taking from the training data using by ratio. The training set and testing set are shown in the table below, along with classification accuracy metrics for the train and testing data separately.

Table 7. Result of experiments by using two training and testing dataset ratio

Training Split	Testing Split	CNN Model Accuracy	Hybrid of CNN and LSTM Model Accuracy
70%	30%	99.04	98.72
80%	20%	99.36	94.40

From those experiments using 80% for training and 20% for testing is better performing or optimal among the 7:3 in both model. Using ratio 8:2 means using 80% of the whole dataset is for training and 20% of the whole dataset is used for testing. In addition to this testing data is taken from the training data. For ratio 7:3 the testing data is taken as 30% of the training data.

4.6.1.2. Experimental 2: Changing Learning Rate for CNN and Hybrid of CNN and LSTM model.

The results of the CNN and Hybrid of CNN and LSTM model experiments using different learning rates are presented in the table below using classification accuracy metrics in the form of percentage for the train data, validation data, and test data separately. As shown in the following result, lower learning rates have lower accuracy then higher learning rates. As a result, the proposed model considers the learning rate of Le-0.001to be optimal.

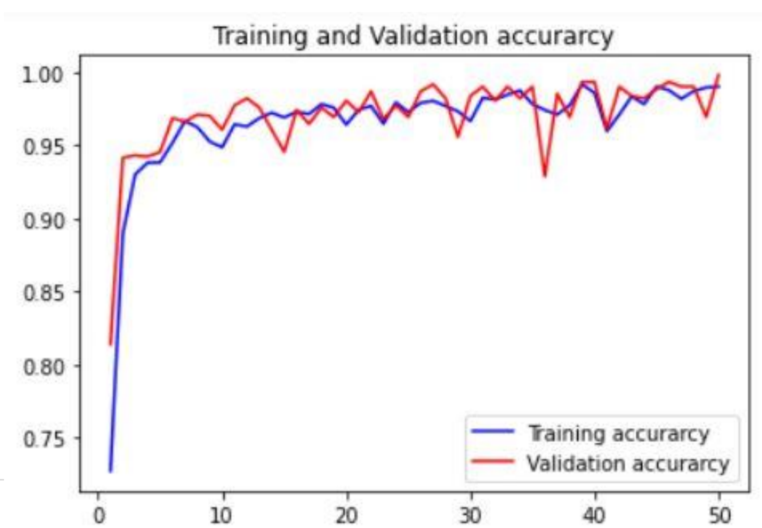
Table 8. Result of the proposed model by using different learning rate

Learning Rate	CNN Model Accuracy	Hybrid of CNN and LSTM Model Accuracy
Le-0.01	96.24%	94.40%
Le-0.001	99.04%	98.72%
Le-3	99.36%	93.60%

As shown in Table 8 and Table 9, the ratio of data and learning rate of the three model are successfully classifies in 8:2 and Le-0.001. In our experiment for the CNN model the accuracy is 99.36% and the Hybrid of CNN and LSTM model accuracy is 98.72% using softmax activation function for multiclass classification.

4.6.2. Result Analysis for the CNN Model of Cervical Cancer Detection and Classification

As shown in Figure 14, at the start of training, the value of the training accuracy line is around 78 percent and the value of the validation accuracy line is around 81 percent, and both values increase up to epoch 6. After epoch 14, both lines have reached 97 percent and are gradually increasing. When we look at the training loss and validation loss curves in Figure 14, we can see that training loss decreases linearly from 0.55 to 0.05 and validation loss decreases linearly from 0.36 to 0.01 from the first to the 50th epoch. After epoch 16, both the validation and training loss lines pass 0.06, which decreases rapidly from the start of the training, but not 0.05, which is the smallest value in the training. Finally, we can see that the validation accuracy is slightly higher than the training accuracy, and the validation loss is slightly lower than the training loss. The validation and training accuracy curves are nearly linear, while the validation and training loss curves are nearly linear. The curves show that there is no overfitting in the CNN model because validation accuracy is increasing rather than decreasing and validation loss is decreasing rather than increasing, and most importantly, there is a small gap between training and validation accuracy and also between training and validation loss. Therefore, we can say that our model's generalization capability became much better since the loss of the validation set was only slightly more compared to the training loss.



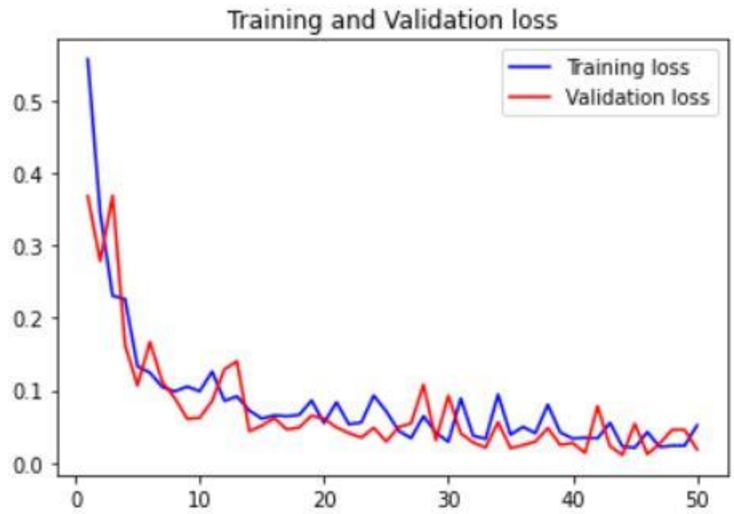
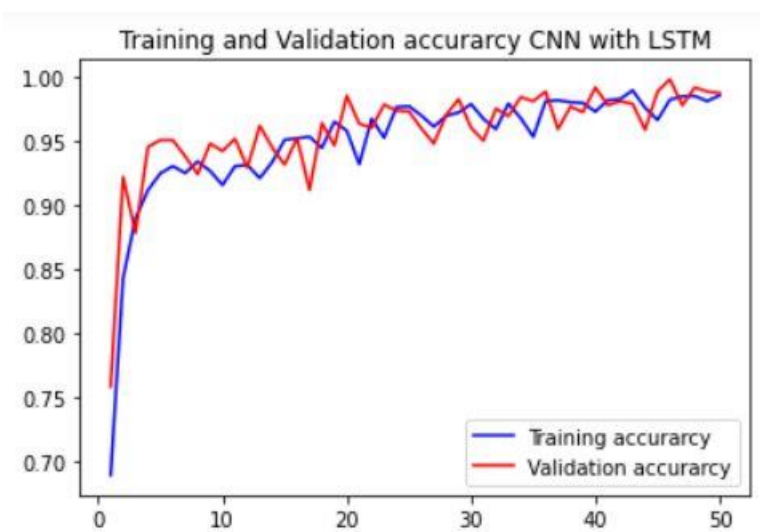


Figure 14. Training and validation accuracy and loss of the CNN model

4.6.3. Result Analysis for the Hybrid of CNN and LSTM Model of Detection and Classification of Cervical Cancer

For the hybrid of CNN with LSTM model result when we see the following plot Figure 15, the training accuracy in the first epoch is around 0.68 and validation accuracy is around 0.75. Then both validation and training accuracy automatically increases when we see the value at epoch 2 and after epoch 27 the values get above 0.97%. The validation accuracy is linearly increasing and no decreasing at the same time the validation loss is linearly decreasing no increasing and have not much gap between training and validation accuracy and loss. Therefore, there is no overfitting problem in the model when we train by using our dataset.



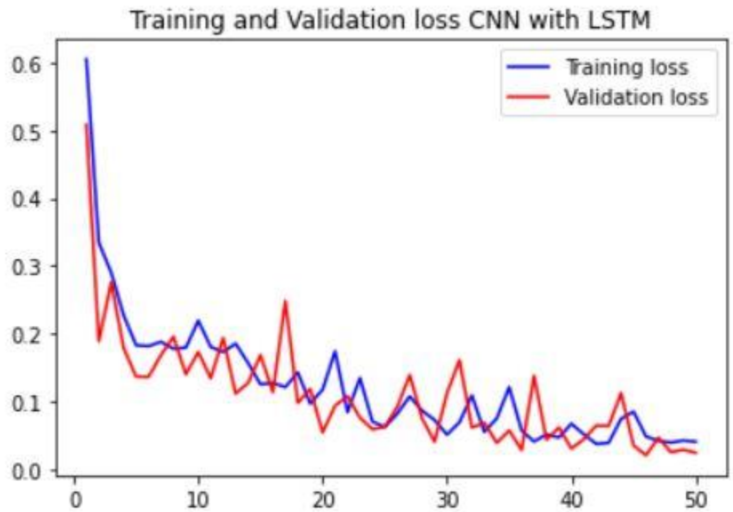


Figure 15. Training and validation accuracy and loss of the Hybrid of CNN and LSTM model

4.7. Performance evaluation metrics

The confusion matrix [51] is a table that describes the performance of a classifier or classification model on a set of test data with known true values. It represents how the classification model becomes perplexed when making predictions. As a result, a confusion matrix is a summary of classification problem prediction results. The confusion matrix is a matrix that is used to assess the performance of classification models for a given set of test data. It can be determined only if the true values of the test data are known. The matrix itself is simple to grasp, but the associated terminologies can be perplexing. Because it displays the errors in the model performance as a matrix, it is also known as an error matrix. Using the confusion matrix, which is one of the metrics used to evaluate the proposed approach, we can calculate the model's various parameters, such as accuracy, precision, recall, and F1-score [52].

According to [51] defines the confusion matrix as a table that describes the performance of a classifier or classification model on a set of test data with known true values. It represents the classification model's confusion when making predictions. As a result, a confusion matrix is a summary of prediction results for classification problems. The confusion matrix is a matrix that is used to evaluate classification model performance for a given set of test data. It is only possible to determine if the true values of the test data are known. The matrix itself is easy to understand, but the associated terminologies can be confusing. It is also known as an error matrix because it displays the errors in the model performance as a matrix.

Classification Accuracy: It is an important parameter in determining the accuracy of classification problems. It specifies how frequently the model predicts the correct outcome. It can be calculated as the ratio of the classifier's correct predictions to the total number of predictions made by the classifiers. The formula is as follows:

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FP+FN+TN}} \tag{4.1}$$

Precision: It can be defined as the number of correct outputs provided by the model or how many of the positive classes predicted correctly by the model were actually true. It can be calculated using the formula below:

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{4.2}$$

Recall: It is defined as the percentage of positive classes predicted correctly by our model out of a total of positive classes. The recall rate should be as high as possible.

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{4.3}$$

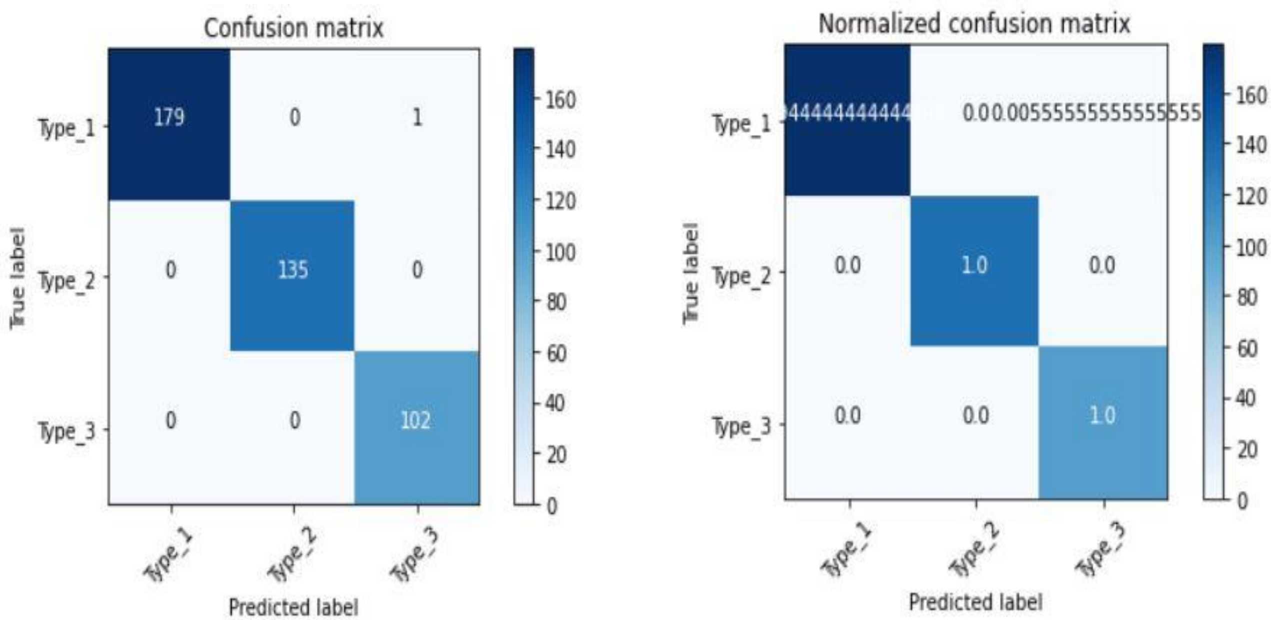
F-measure: It is difficult to compare two models that have low precision but high recall or vice versa. So, we can use F-score for this purpose. This score allows to assess both recall and precision simultaneously. If the recall equals the precision, the F-score is maximized. It can be calculated using the formula below:

$$\text{F} - \text{measure} = \frac{2*\text{Recall}*\text{Precision}}{\text{Recall+Precision}} \tag{4.4}$$

4.7.1. Performance evaluation metrics results analysis

The following metrics are used to measure the performance of the proposed system: TP denotes the correctly predicted Type_1, Type_2, and Type_3. FP denotes Type_1, Type_2, and Type_3 that are misclassified as Type_1, Type_2, and Type_3 by the proposed system, TN denotes the Type_1, Type_2, and Type_3 that are correctly classified, and FN denotes the Type_1, Type_2, and Type_3 that are misclassified as Type_1, Type_2, and Type_3 shown in Figure 16 and 17.

As it is shown blow in the confusion matrix of each classifier, Values on the main diagonal of the confusion matrix represents the number of categories correctly classified to the corresponding category. In this study to generate the confusion matrix of each classifier the sklearn package is used. Referring to the results obtained, as shown in Figure 16 and 17 illustrated the performance of comparison of each classification techniques.



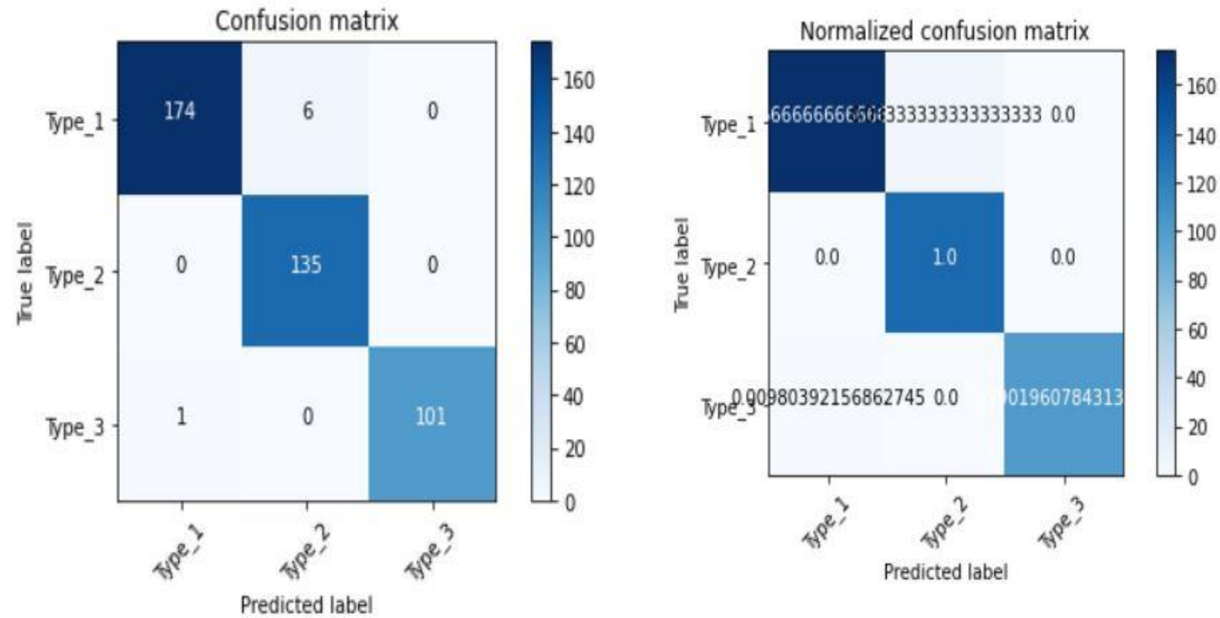

```
              precision    recall  f1-score   support

      Type_1       1.00      0.99      1.00       180
      Type_2       1.00      1.00      1.00       135
      Type_3       0.99      1.00      1.00       102

    accuracy                           1.00       417
   macro avg       1.00      1.00      1.00       417
weighted avg       1.00      1.00      1.00       417

[[179   0   1]
 [  0 135   0]
 [  0   0 102]]
```

Figure 16. Shows the Confusion Matrix value of CNN.



	precision	recall	f1-score	support
Type_1	0.99	0.97	0.98	180
Type_2	0.96	1.00	0.98	135
Type_3	1.00	0.99	1.00	102
accuracy			0.98	417
macro avg	0.98	0.99	0.98	417
weighted avg	0.98	0.98	0.98	417

Figure 17. Shows the Confusion Matrix value of hybrid of CNN with LSTM.

4.8. Discussion

As presented in the previous sections, the experiments were conducted by using deep learning model like CNN and hybrid of CNN with LSTM model. All of the experiments are conducted using the same hardware configuration. The number of images in the dataset which are used to train the both models is similar but, the depth of the both models or number of parameters are different. Both model is trained with a total of 2085 original images. Both of the models are tested by testing dataset and obtained good result. Classification accuracy metrics are used to measure the performance of the models and the proposed model has better classification results.

As we can see in the following table 9 the precision, recall and f1-score of CNN and hybrid of CNN with LSTM model is described below. In this study the CNN model is achieved high accuracy then hybrid of CNN with LSTM based on the below result. The performance metrics of CNN model based on precision, recall and f1-score for Type_1 cervical cancer is 1.00, 0.99 and 1.00 respectively, for cervical cancer of Type_2 disease is 1.00, 1.00 and 1.00 respectively and at last Type_3 cervical cancer result is 0.99, 1.00 and 1.00 respectively. These show the models are giving good result on this model. The performance metrics of hybrid of CNN with LSTM model based on precision, recall and f1-score for Type_1 cervical cancer is 0.99, 0.97 and 0.98 respectively, for cervical cancer of Type_2 disease is 96.6, 1.00 and 98.49 respectively and at last Type_3 cervical cancer result is 1.00, 0.99 and 1.00 respectively.

Performance metrics	**CNN**			**Hybrid of CNN with LSTM**		
	Type-1	Type-2	Type-3	Type-1	Type-2	Type-3
Precision	1.00	1.00	0.99	0.99	0.96	1.00
Recall	0.99	1.00	1.00	0.97	1.00	0.99
F1-score	1.00	1.00	1.00	0.98	0.98	1.00

Table 9.summery of performance metrics

The main reasons that the proposed model gives better result are because of our proposed model uses smaller sized filters in the convolution layer of the network. Using smaller sized convolution helps to identify very smaller features which are used to distinguish between the input image and the probability of losing an important feature is very less.

Most deep learning algorithms especially computer vision for image classification problems are trained by using high performance computing machines with faster GPU, a huge number of images (in millions), and tens of millions of parameters. But we can train and get better results with small sized networks with fewer parameters, less hardware consumption, and fewer data. The other point is that making preprocessing to the images by removing noise and unwanted features will increase the accuracy of the model.

CHAPTER FIVE

5. CONCLUSIONS AND RECOMMENDATIONS

5.1. Overview

Cervical cancer is a type of cancer that affects a woman's cervix. The cervix is the uterus's neck-shaped passage at the bottom. This cancer is a headache for every countries on the world.in rest of the countries there are frequently checkup for the cancer while in developing countries like Ethiopia the women come to hospital after they are affected. This cause a high mortality rate in the country. So to solve this problem early detection model was needed and the research was carried out. The research is included as follows.

5.2. Conclusions

Cervical cancer is a leading cause of death from cancer among women in low-resource settings, affecting women at a time of life when they are critical to social and economic stability. The aim of this study is to developing deep learning based cervical cancer disease detection and classification model. To achieve the main objectives of this study the literature on artificial intelligence, computer vision, machine learning, image processing and deep learning was reviewed. The algorithm and both software and hardware tools were determined. The research was conducted to develop a detection and classification model in different deep learning algorithms. From those deep learning algorithm we use CNN and hybrid of CNN with LSTM model to train and test the cervical cancer to detect and classify the cervical disease. The developed models evaluated and achieved the efficient accuracy. The model developed on CNN achieved macro average using precision is 100%, Recall 100% and F1-score 100% and hybrid of CNN with LSTM achieved macro average using precision is 98%, Recall 98% and F1-score 98% respectively.

The results we have in our experiment has proven that the proposed CNN and Hybrid of CNN and LSTM models can significantly support for accurate detection and classification of cervical cancer of Type-1, Type-2 and Type with little computational effort and little images which is far less than that of expected for deep learning algorithm because most of deep learning algorithms are trained with millions of images with high computational resource. To this end we are encouraged by the obtained results from the experiment and, we are intended to work and to test more cervical cancer images and class with our model. CNN and hybrid of CNN and LSTM models have better classification results, for this reason both algorithms are more suitable algorithms to detect and classify cervical cancer.

5.3. Recommendations

There are many diseases stage types that can affect Cervix. This study specifies the classes into Type_1, Type_2 and Type_3. The Normal stages are out-scoped due to lack of data. To make the model more reliable, if an amount of disease class increases with high large data, then the problem of chasing disease can be reduced.

The future work of this study is developing Mobile applications that can be deployed into Android to use the model for scout teams.

References

[1] Huang, Pan; Zhang, Shuailei ; Li, Min; Wang, jing; Ma, Cailing; Wang, Bowei; Lv, Xiaoyi, "Classification of Cervical Biopsy Images Based on LASSO and EL-SVM," *IEEE Access*, pp. 24219 - 24228.

[2] Peng, Y., Kang, Q., Luo, Q., Jiang, W., Si, W., Liu, B.A., Luu, H.H., Park, J.K., Li, X., Luo, J., Montag, A.G., Haydon, R.C. , He, T.C., "Inhibitor of DNA binding/differentiation helix-loop-helix proteins mediate bone morphogenetic protein-induced osteoblast differentiation of mesenchymal stem cells," *The Journal of Biological Chemistry,* pp. 32941-32949, 2004.

[3] RAHAMAN, MD MAMUNUR; LI, CHEN; WU, XIANGCHEN; YAO, YUDONG; HU, ZHIJIE; JIANG, TAO; LI, XIAOYAN; QI, SHOULIANG, "A Survey for Cervical Cytopathology Image Analysis Using Deep Learning," *IEEE Acsess,* vol. 8, pp. 61687 - 61710.

[4] Begoihn, Matthias; Mathewos, Assefa; Aynalem, Abreha; Wondemagegnehu, Tigeneh; Moelle, Ulrike; Gizaw, Muluken; Wienke, Andreas; Thomssen, Christoph; Worku, Dawit; Addissie, Adamu; Jemal, Ahmedin; Kantelhardt, Eva Johanna, "Cervical cancer in Ethiopia - predictors of advanced stage and prolonged time to diagnosis," *National Library of Medicine National Center for Biotechnology Information.*

[5] J. S. Gero, Research Methods for Design Science Research: Computational and Cognitive Approaches, Australia: In Proceedings of ANZAScA, 1999.

[6] Wei, Shuli; Dai, Peifeng; Wang, Zhengping, "Cervical Cancer Detection and Diagnosis Based on Saliency Single Shot MultiBox Detector in Ultrasonic Elastography," *National Library of Medicine National Center for Biotechnology Information,* vol. 3217, pp. 873-880, 2004.

[7] Arya, Mithlesh; Mittal , Namita, "Cervical Cancer Detection Using Single Cell and Multiple Cell Histopathology Images: Do You Have a Subtitle? If So, Write," *ResearchGate*, pp. 205-215, 2019.

[8] "What Is Cervical Cancer?," American Cancer Society, 30 July 2020. [Online]. Available: https://www.cancer.org/cancer/cervical-cancer/about/what-is-cervical-cancer.html.

[9] V. Chandran, S. M. G, A. Karthick and T. George, "Diagnosis of Cervical Cancer based on Ensemble Deep Learning Network using Colposcopy Images," *BioMed Research International* , May 2021.

[10] Z. Zhang, X. He, X. Sun, L. Guo and J. Wang, "Image Recognition of Maize Leaf Disease Based on GA-SVM," 2015. [Online].

[11] B. Chitradevi and P. Srimathi, "An Overview on Image Processing Techniques," 2014.

[12] V. Dixit and J. Pruthi, "Review of Image Processing Techniques for Automatic Detection of Tumor in Human Liver," *International Journal of Computer Science and Mobile Computing*, vol. 3, p. 371–378, 2014.

[13] T. Berhanu, "Survey of diseases of major crops in Darolebu district,West Hararge, Ethiopia," *International Journal of Plant Breeding and Crop Science*, vol. 2, p. 43–45, 2015.

[14] J. D. Pujari, R. Yakkundimath and A. S. Byadgi, *International Journal of Interactive Multimedia and Artificial Intelligence,* vol. 3, p. 6–14, 2016.

[15] D. M. M. A. S. Abrham and G. Mengistu, "Ethiopian Coffee Plant Diseases Recognition Based on Imaging and Machine Learning Techniques," *International Journal of Database Theory and Application,* vol. 9, p. 79–88, 2016.

[16] J. Arcia and A. Barbedo, "Digital image processing techniques for detecting , quantifying and classifying plant diseases," p. 1–12, 2013.

[17] Birhanu Hailu, M. M, "Applying Image Processing for Malt-barley Seed Identification," in *In Ethiopian the 9th ICT Annual Conference 2016 (EICTAC 2016)*, Bahir Dar, Ethiopia, 2016.

[18] W. F. ,. ,. Arif, A. Saduf and I. K. Asif, Advances in Deep Learning, vol. 57, Springer Nature Singapore Pte Ltd, 2020, pp. 2-3.

[19] J. Schmidhuber, Deep Learining in neural networks: An overview, Neural Networks, vol. 61, 2015, pp. 85-117.

[20] G. Ian, B. Yoshua and C. Aaron, Deep Learning, MIT press, 2016.

[21] A. K. Jain, J. Mao and K. M. Mohiuddin, Artificial neural networks, A tutorial In: Computer 29.3, 1996, p. 31–44.

[22] S. Sharma, "Activation Functions in Neural Networks," 6 9 2017. [Online]. Available: https://towardsdatascience.com/activation-functions-neural-networks. [Accessed 11 8 2021].

[23] S. Mahanama, "Introduction to Artificial Neural Networks(ANN)," Analytics Vidhya, 10 Jul 2020. [Online]. Available: https://medium.com/analytics-vidhya/introduction-to-artificial-neural-networks-ann-3109578d61ab.

[24] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," in *IEEE Conference on Computer Visio and Pattern Recognition*, 2016.

[25] A. Krizhevsky, I. Sutskever and G. Hinton, ImageNet classification with deep convolutional

neural networks, Adv Neural Inf Process Syst 25, 2012.

[26] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich, Going deeper with convolutions, In: Cvpr, 2015.

[27] Brilliant, "Feedforward neural network," 2018.

[28] A. Zell, Simulation Neuronaler Netze [Simulation of Neural Networks], 1 ed., German: Addison-Wesley, 1994, p. 73.

[29] U. Karn, "A Quick Introduction to Neural Networks," 2016.

[30] J. Brownlee, "How to Choose an Activation Function for Deep Learning," 22 January 2021. [Online]. Available: https://machinelearningmastery.com/choose-an-activation-function-for-deep-learning. [Accessed 12 8 2021].

[31] A. Fernandez, "datascience," 29 20 2020. [Online]. Available: https://datascience.aero/aviation-function-deep-learning/. [Accessed 11 8 2021].

[32] S. Sootla, "Artificial neural network for image classification," Computational neurosicence univesity of Tartu, 2015.

[33] M. Valueva, N. Nagornov, P. Lyakhov, G. Valuev and N. Chervyakov, Application of the residue number system to reduce hardware costs of the convolutional neural network implementation, Mathematics and Computers in Simulation, 2020, pp. 232-243.

[34] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley and S. C. Ozair, Advances in neural information processing systems, Generative adversarial nets, 2014, pp. 3,29,31.

[35] C. Aggarwal, Neural Networks and Deep Learning, Switzerland AG: Springer International Publishing AG,part of Springer Nature, 2018, pp. 339-349.

[36] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, . J. Santamaría, M. A. Fadhel, M. Al-Amidie and L. Farhan , "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *Journal of Big Data*, vol. 8, no. 53, 2021.

[37] R. Yamashita, M. Nishio, . R. K. G. Do and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights Imaging*, vol. 9, p. 611–629, 2018.

[38] C. Bishop, Training with noise is equivalent to Tikhonov regularization, vol. 7, Neural computation, 1995, pp. 108-116.

[39] R. Fergus and M. Zeiler, "Visualizing and understanding convolutional networks," in *European Conference on Computer Vision*, 2013.

[40] C. Szegedy, S. Ioffe, V. Vanhoucke and A. Alemi, "Inception-v4, Inception-ResNet and the Impact of Residual Connection on Learning," in *AAI Conference*, 2017.

[41] A. Graves, G. Hinton and A. Mohamed, "Speech recognition with deep recurrent neural networks," in *IEEE international conference on acoustics, speech and signal processing*, 2013.

[42] F. A. Gers, J. Schmidhuber and F. Cummins, Learning to forget: Continual prediction with LSTM, 1999.

[43] D. C. Novitasari, A. Z. Foeady and A. H. Asyhar, "Automatic Approach for Cervical Cancer Detection Based on Deep Belief Network (DBN) Using Colposcopy Data," *Semantic Scholar,* p. 2020.

[44] E. Karim and N. Neehal, "An Empirical Study of Cervical Cancer Diagnosis using Ensemble Methods," pp. 1-5, 2019.

[45] P. Gupta, I. Jindal and A. Goyal, "Early Detection and Prevention of Cervical Cancer," *Medicine 2019 IEEE 5th International Conference for Convergence in Technology (I2CT),* 2019.

[46] B. Wimpy and S. Suyanto, "Classification of Cervical Type Image Using Capsule Networks," *Medicine, 2019 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI),* pp. 34-37, 2019.

[47] K. K. Allehaibi, L. Nugroho, W. L. Lazuardi, A. S. Prabuwono and T. Mantoro, "Segmentation and Classification of Cervical Cells Using Deep Learning," *Semantic Scholar,* 2019.

[48] C. Francois, Deep Learning with Python, New York: Manning Publications, 2017.

[49] A. Kulkarni and F. A. B. , "Foundations of data imbalance and solutions for a data democracy," 2020.

[50] K. K. Al-jabery and D. C. W. I. , "Data analysis and machine learning tools in MATLAB and Python," 2020.

[51] M. Sunasra, "medium.com," 11 Nov 2017. [Online]. Available: https://medium.com/@MohammedS/performance-metrics-for-classification-problems-in-machine-learning-part-i-b085d432082b. [Accessed 8 10 2021].

[52] Bharathi;, "Analytics Vidhya," 24 June 2021. [Online]. Available: https://www.analyticsvidhya.com/b;pg/2021/06/confusion-matrix-for-multi-class-classification. [Accessed 8 October 2021].

[53] Pouyanfar, Samira and Sadiq, Saad and Yan, Yilin and Tian, Haiman and Tao, Yudong and

Reyes, "A survey on deep learning: Algorithms, techniques, and applications," *ACM Computing Surveys (CSUR),* vol. 51, no. ACM New York, NY, USA, pp. 1-36, 2018.

[54] Ajay Shrestha, and Mahmood, Ausif, "Review of deep learning algorithms and architectures," *IEEE Access,* no. IEEE, 2019.

[55] Editorial, "Different types of Deep Learning models explained," RoboticsBiz, 15 March 2015. [Online]. Available: https://roboticsbiz.com/different-types-of-deep-learning-models-explained/. [Accessed 28 9 2020].

[56] R. S. Dwivedi, Digital Image Processing, Germany, 2017.

[57] A. Mathew, P. A. and S. S. , "Deep Learning Techniques: An Overview," *researchgate.net/publication/341652370,* 02 August 2020.

[58] A. Khan, A. S. U. Z. and A. S. Qureshi, "A Survey of the Recent Architectures of Deep Convolutional Neural Networks," *Artificial Intelligence Review*, pp. 1-71, 15 May 2020.

[59] A. Graves, A. r. Mohamed and G. Hinton, "Speech recognition with deep recurrent neural networks," in *IEEE international conference on acoustics, speech and signal processing*, 2013.

[60] F. A. Gers, J. Schmidhuber and F. Cummins, Learning to forget: Continual prediction with LSTM, 1999.

[61] J. Patterson and A. G. , "Deep Learning," in *Deep Learning*, O'Reilly Media, Inc., 1005 Gravenstein Highway North,Sebastopol, CA 95472, 2017.

[62] R. Yamashita, M. N. R. K. Gian Do and K. T. , "Convolutional neural networks: an overview and application in radiology," 22 June 2018.

[63] F. Li, J. Justin and Y. S. , "Convolutional Neural Networks for Visual Recognition," 20 January 2019.

[64] F. Li, J. Justin and Y. S. , "Convolutional Neural Networks for Visual Recognition," 20 january 2019.

[65] D. A. Khan, "Introduction to Deep Convolutional Neural Networks," *ResearchGate,* 08 November 2016.

[66] S. S. Ambekar and M. A. C. , "ARTIFICIAL NEURAL NETWORK CONTROLLER FOR PERFORMANCE OPTIMIZATION OF SINGLE PHASE INVERTER," *International Journal of Artificial Intelligence & Applications (IJAIA),* vol. Vol.3, January 2012.

[67] S. Sharma, S. S. and A. A. , "ACTIVATION FUNCTIONS IN NEURAL NETWORKS," *International Journal of Engineering Applied Sciences and Technology,* vol. Vol. 4, no.

Issue 12, ISSN No. 2455-2143, pp. 310-316, 2020.

[68] K. Gurney, An introduction to neural networks, London and New York, 2004.

[69] J. Zupan, "Introduction to Artificial Neural Network (ANN) Methods: What They Are and How to Use Them," September 1994.

[70] M. Buscema, "Back Propagation Neural Networks," *Semeion Research Center of Sciences of Communication*, 09 May 2015.

[71] François Chollet, Deep Learning with Python, Manning Publications, 2017.

[72] "Adobe Photoshop CS5 Tutorial".

[73] C. Metz, "TensorFlow, Google's Open Source AI, Points to a Fast-Changing Hardware World," Wired, 2015.

[74] A. Geron, "Hands On Machine Learning with Scikit-Learn, Keras and TensorFlow," 2019.