SAIF SALAM JEBUR

GASTRIC CANCER CLASSIFICATION USING DEEP CONVOLUTIONAL NEURAL NETWORK

THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
ATILIM UNIVERSITY

SAIF SALAM JEBUR

A MASTER OF SCIENCE
THESIS
IN
ELECTRIC AND ELECTRONIC ENGINEERING

APRIL 2020

ATILIM UNIVERSITY

GASTRIC CANCER CLASSIFICATION USING DEEP CONVOLUTIONAL NEURAL NETWORK

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
ATILIM UNIVERSITY

BY

SAIF SALAM JEBUR

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE-ARCHITECTURE /DOCTOR OF PHILOSOPHY
IN
ELECTRIC AND ELECTRONIC ENGINEERING

APRIL 2020

Approval of the Graduate School of Natural and Applied Sciences, Atilim University.

_____

Prof. Dr. Ali Kara

Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of **Master of Science in Electric and Electronic Engineering, Atilim University**.

_____

Assoc. Prof. Dr. Efe Eseller

Head of Department

This is to certify that we have read the thesis "**Gastric Cancer Classification Using Deep Convolutional Neural Network**" submitted by "Saif Salam Jebur" and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

_____

Asst. Prof. Dr. Hakan TORA
Supervisor

Asst. Prof. Dr. Emre Sümer
Computer Eng. Dept. , Başkent University _____

Asst. Prof. Dr. Hakan TORA
Electric and Electronics Eng. Dept. , Atilim University_____

Asst. Prof. Dr. Erhan GÖKÇAY
Software Eng. Dept. , Atilim University _____

Date: *27.4.2020*

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name : Saif Salam Jebur

Signature :

ABSTRACT

GASTRIC CANCER CLASSIFICATION USING DEEP CONVOLUTIONAL NEURAL NETWORK

Jebur, Saif Salam

MSc., Electric and Electronic Engineering

Asst. Prof. Dr. Hakan Tora

April 2020, 122 pages

In this thesis, several pre-trained CNN and our CNN structure presented to automatic detection of early gastric cancer in endoscopic images. In the first stage, the transfer learning using two types normal and cancer of image datasets, the pre-trained networks executed for gastric cancer detection using MATLAB 2018. Then, the obtained results compared with each other and discussed in detail form. In the second stage, new structure proposed by using CNN. The proposed structure consists from 8 layers with SoftMax classifier. The extracted high-level features by convolutional layers classified by SoftMax in last layer. The proposed network presented 99.88% which is high result when compared with numerous performed pre-trained networks. Furthermore, the proposed network presented remarkable execution time when compared with several transfer learning techniques.

Keywords: CNN, Deep learning, Gastric cancer, Machine learning, SoftMax.

ÖZ

DERİN ÇEVRESEL SİNİR AĞINI KULLANARAK MİDE KANSER SINIFLANDIRMASI

Jebur, Saif salam
Yüksek Lisans., Elektrik Elektronik Mühendisliği
Dr. öğretim üyesi Hakan Tora

Nisan 2020, 122 sayfa

Bu tezde, önceden eğitilmiş birkaç CNN ve CNN yapımız endoskopik görüntülerde erken mide kanserinin otomatik olarak tespit edilmesine sunulmuştur. İlk aşamada, iki tip normal ve görüntü veri kümelerinin kanseri kullanılarak yapılan transfer öğrenimi, MATLAB 2018 kullanılarak mide kanseri tespiti için önceden eğitilmiş ağlar gerçekleştirildi. Daha sonra elde edilen sonuçlar birbirleriyle karşılaştırıldı ve ayrıntılı olarak tartışıldı. İkinci aşamada, CNN kullanılarak önerilen yeni yapı. Önerilen yapı SoftMax sınıflandırıcılı 8 katmandan oluşur. Son katmanda SoftMax tarafından sınıflandırılan evrişimsel katmanlarla çıkarılan yüksek seviye özellikler. Önerilen ağ 99.88% sundu ve bu da önceden eğitilmiş birkaç ağla karşılaştırıldığında yüksek sonuçtur. Ayrıca, önerilen ağ, çeşitli transfer öğrenme teknikleriyle karşılaştırıldığında dikkate değer bir yürütme süresi sundu.

Anahtar Kelimeler: CNN, Derin öğrenme, Mide kanseri, Makine öğrenmesi, SoftMax.

To My Parents

My Father "Salam Jebur Saloom"

My Mother "Khansaa Ibrahim Hasan"

ACKNOWLEDGMENTS

I want to communicate true thanks to my supervisor Asst. Prof. Dr. Hakan Tora for his direction, support. Tolerance and knowledge all through my thesis study. I offer earnest a debt of gratitude is in order for his broad information, motivating force and persistence during this period. I am extremely glad for being one of his students.

Besides, I would likewise communicate my thankfulness to examination committee members, for their significant time, remarks and proposals.

Last but not least, I want to communicate my gratefulness to my father Salam Jebur Saloom, and my mother Khansaa Ibrahim Hasan for their perpetual love, backing and tolerance through the study and for my entire life.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF SYMBOLS

CNN	Convolutional Neural Network
RNN	Recurrent Neural Netwok
DL	Deep Learning
ML	Machine Learning
SVM	Support Vector Machine
DT	Decision Tree
DNN	Deep Neural Network

CHAPTER 1

INTRODUCTION

Gastric cancer is a sickness in which evil (cancer) lockups procedure in the coating of the abdominal. Age, food and abdominal sickness can touch the danger of evolving stomach cancer. Indications of stomach cancer contain dyspepsia and stomach distressed or discomfort.

Primary gastric cancers, wherever tumor cells are narrowed to the mucosa, must remained recognized where there is dynamic screening of patients at high-risk for gastric cancer. In these patients, primary gastric cancer might seem as an understated cut, typically fewer than 2 cm in thickness. The detection of initial gastric cancer is essential since it is possibly agreeable to endoscopic treatment and escorted by an outstanding prognosis [1].

Furthermore, it is essential to offer exact, fast screening for gastric cancer. If a patient is estimated as existence at tall risk, then (s) he container pursue to assume preventive actions in advance. Equally, if a patient is estimated as being at small danger, then (s) he can escape or decrease the frequency of higher gastrointestinal endoscopic checks. Some new researches have confirmed that new techniques like machine learning and deep learning methods are effective for enhancing screening, detection, biomarker extraction and disease predication in the medical area. The researcher offered that complete screening applying a mixture of many factors accrued each day in hospitals and effective deep learning methods could lead to additional accurate and fast screening for gastric cancer. Because of increasing the ratio of people injured in Gastric cancer in east Asia countries such as Japan, and China see Figure 1.1 [2]. the researcher really tries to apply deep learning techniques to this field.

Deep learning has been lately become a trend and lead an entire new study field. In fact, they have been applied in numerous areas. Between the several kinds of academia and manufacturing, the request for AI in the area of health informatics has improved, and the possible assistances of AI applications in healthcare have also been enhanced. In this thesis, CNN applied to detect the gastric cancer by using several pre-trained networks such as (AlexNet, GoogleNet, Vgg16, DenseNet201 and ResNet101) The

obtained results compared with each other and with several previous studies presented in this field.

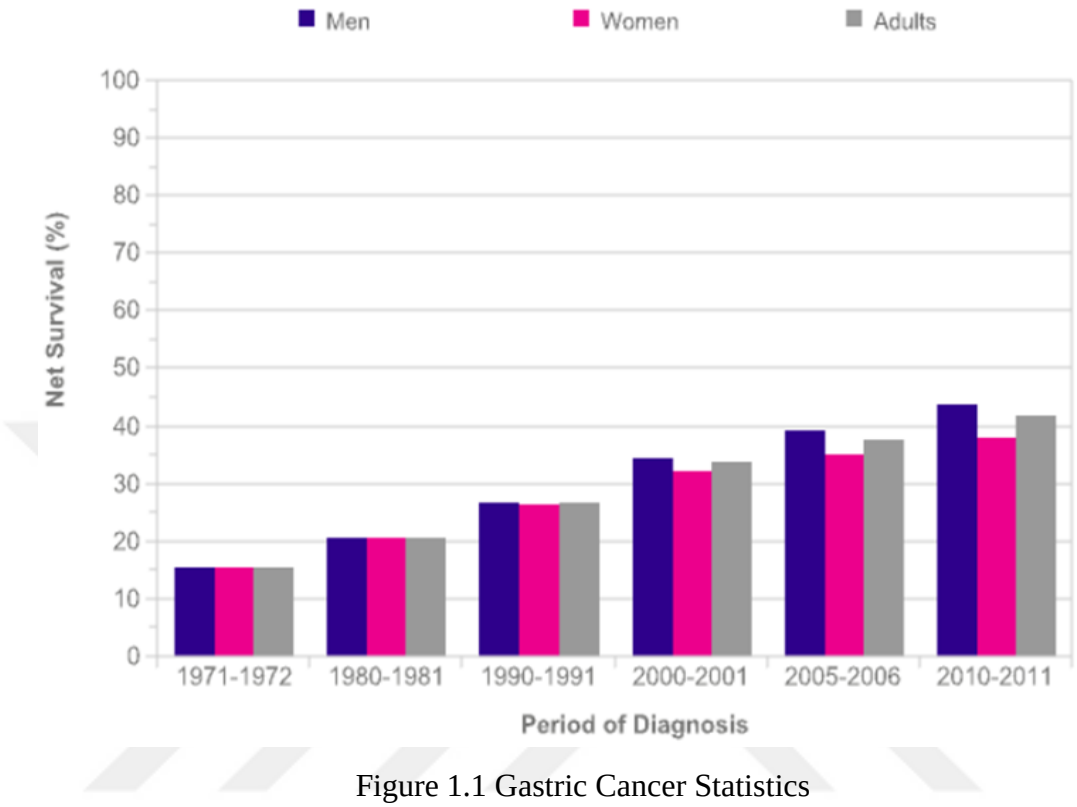

Figure 1.1 Gastric Cancer Statistics

1.1 Contributions

In this part, several points are listed below to explain our contributions in this research:

- Developed application to detect gastric cancer. The deep learning technique CNN applied to recognize gastric cancer. Several CNN pre-trained techniques applied to gastric cancer such as AlexNet, GoogleNet, ResNet101, VGG16, DenseNet201. Furthermore, we proposed new structure of CNN without using transfer learning or any pre-trained techniques to deal with gastric cancer.
- Applied CNN pre-trained networks and compare the results. The five pre-trained CNN method results and our network are compared with our structure. Moreover, this comparison show that our structure presented best results.

- Presented previous studies in detail form. Our experimental results compared with well-known studies presented in this field using the same dataset. Then, the comparison show that our results is best than previous studies.

1.2 Research Questions

In this part, several questions are proposed in the beginning of this study as research questions are listed below, the author during the research try to answer these questions:

- In this study, the author tries to solve the recognition of the gastric cancer automatically. So, after deep searching in the gastric cancer, the author first question is how to developed effective application to detect gastric cancer?
- By searching in the AI techniques, the author investigates which machine learning and deep learning techniques?
- How applied CNN to gastric cancer? several transfer learning techniques are applied to the same dataset such as AlexNet, GoogleNet, ResNet101, VGG16, DenseNet201 and our structure also presented to the same problem?
- Compare the results with previous studies?

CHAPTER 2

RELATED WORK

2.1 Gastric Cancer

Gastric cancer is the 3rd highest popular source of cancer-related killing in the world, [1] and it continues challenging to treat in Western countries, mainly for highly patients current with complex disease. In the United States, stomach cancer is presently the 15th highest general cancer. [3] The stomach starts at the gastroesophageal intersection and closes at the duodenum. See the Figure 2.1 below.

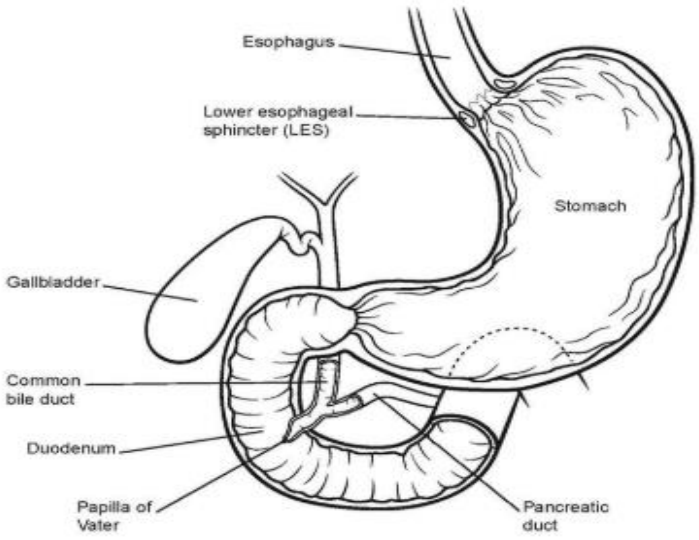

Figure 2.1 The Digestive System

There are several Symptoms for gastric cancer such as Tiredness, the person felling he is full after eating little bit, the person felling in bloated after eating, pain in the gastric system, and loss high weight in a short time [4]. There are several factors led to injure in gastric cancer some of them are listed below:

- Age: There is a high-pitched growth in gastric cancer degrees in people over age 50. Greatest persons detected with gastric cancer are among their evening 60s and 80s.

- Bacteria: A mutual bacterium named Helicobacter pylori, similarly, named H. pylori, reasons stomach irritation and ulcers. It is moreover measured one of the chief reasons of stomach cancer. Validating for H. pylori are obtainable, and a contagion container be preserved with antibiotics.
- Diet: Researchers consider that consumption treated foods might show a part in the growth of stomach cancer. Furthermore, organic food may defend against this sickness.
- Gender: men injured in Gastric cancer more common than women.
- Smoking: Smoking is also having great and high effective that lead to injured in Gastric cancer.
- Previous surgery: if any people have stomach surgery or any stomach problem such as anemia pernicious, these people have high potential to injured with stomach cancer.
- Obesity: The high weight lead also to Gastric cancer in men, but this feature not affect in women such as in men.

Several tests assist to diagnosis the Gastric cancer but the best and commonly used is the endoscopy. The endoscopy assists physicians to treat and analysis the patient's case and understand the details of the case see Figure 2.2.

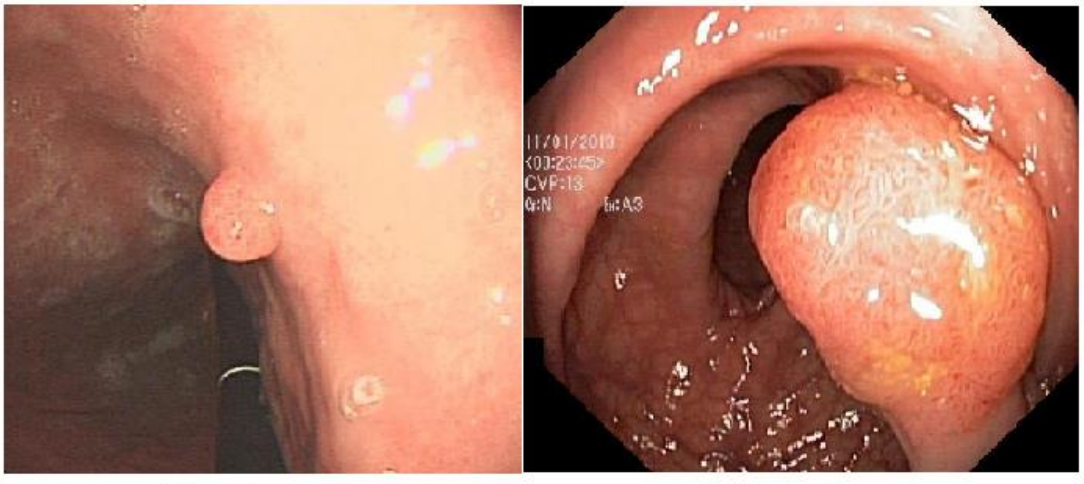

Figure 2.2 Endoscopy Images for Gastric Cancer

Endoscopy image in the Figure 2.2 part A Show that there is a small tumor in the gastric this can be detected and classified by the doctors or any medical software. But in Figure 2.2 part b the tumor is very big and its gastric cancer if the tumor not detected in the part A in earlist time its lead to gastric cancer and its become difficult to treat.

2.2 Software Application Healthcare

Patients, doctors and managers want to make sure that their health care facilities are safe. The organization concerned may be a general practice, a ward, a department, an entire hospital, or a health system. But what exactly do we mean when we ask whether a health facility is safe? Anyone who listens to a panel, clinical meeting, or any group discussing this question will know that many different perspectives will if not always, be enthusiastically advancing and advocating.

With advances in technology, a typical health care network is both a constantly evolving and expanding structure. For example, healthcare organizations have begun to use mobile devices such as smartphones and tablets to perform tasks such as accessing medical records and sending prescriptions. In many ways, future-oriented developments improve the quality of care patients receive. However, as the number of devices connected to your IT network continues to increase, there is a greater need for surveillance [5].

Nowadays, there is more chance of stopping a patient's journey than ever before. Not just digital medical devices, they rely more on traditional IT infrastructure than in previous years. They connect to the network and need to share all kinds of information with different endpoints.

Despite the importance of medical devices, as in any information system, it is the data at the heart of the modern medical IT system. Patient data must be available at various contact points in the hospital - for example: for the doctor, patient, x-ray department, etc. Images from imaging devices such as ultrasound or MRI images should be transmitted, stored and made available to doctors and nurses inside and outside the hospital.

If the improvement of all people is at the heart of why we collect data, then everyone has the best public health benefit. As a matter of fact, SnapRx will produce a real-time data of approximately 23 million prescriptions per month in the Philippines. These data benefit millions of people by reporting drug recalls, revealing trends in drug use, and preventing disease outbreaks. Surveillance of health data is crucial to understanding the true burden of disease and developing good policies in developing countries. For example, governments and health care organizations can use data provided by a platform such as SnapRx to make informed policy decisions that help improve public health.

The biggest challenge we face with real-time health data monitoring is to make sure that the data is not misused. Misuse of data can lead to discrimination and exploitation, especially in vulnerable populations. For example, health insurance companies should not use people's health data to increase premiums and insure some, and some should not. In addition, the benefits of data should be shared equally among all people. Equal access to data-based information is absolutely essential to reducing social inequality.

Ultimately, real-time data monitoring in health care provides great benefits and can combine useful data for the public good. However, the challenge is to protect privacy and ensure that data is not misused. In fact, mClinica's SnapRx improves pharmacy care while maintaining patient privacy and continues to collect public health data sets that can improve health outcomes for all. Finally, it should be remembered that the data is a powerful tool that should only be used for social progress [6].

2.3 Machine Learning

2.3.1 K-Nearest Neighbours (KNN)

KNN is a family of classification technique and regression technique is frequently mentioned to as memory-depend training or example-depend training. Occasionally, it is also entitled lazy learning. These relations agree to the chief idea of KNN. The idea is to substitute system formation by learning the training dataset and formerly apply this information to brand estimates.

Let’s proceeds a humble situation to comprehend this technique. Subsequent is a feast of red circles and green squares [7] presented in Figures 2.3:

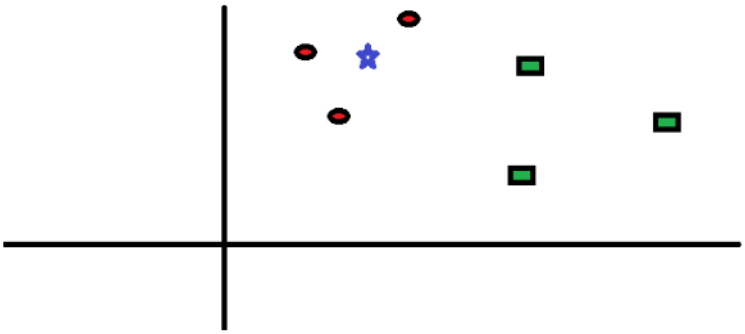

Figure 2.3 Data From 3 Different Points

This mean to bargain out the lesson of the blue star. blue star can also be of red circles or green squares and nonentity different. The “K” is KNN technique is the nearest neighbors we request to income ballot from. Let’s say K = 3. Therefore, this will nowadays brand a circle with blue star as midpoint fair as large as to encircle only three facts points on the flat. Mention to next diagram for more particulars.

The 3 neighboring points to blue star is all red circles. Therefore, with decent sureness stage we can say that the blue star should fit to the lesson red circles. Now, the excellent became actual clear as all 3 ballots from the neighboring neighbor left to red circles. The excellent of the factor K is actual critical in this technique. Following we will comprehend what are the parameters to be careful to accomplish the finest K. The procedures of KNN are presented in Figures 2.4 and 2.9.

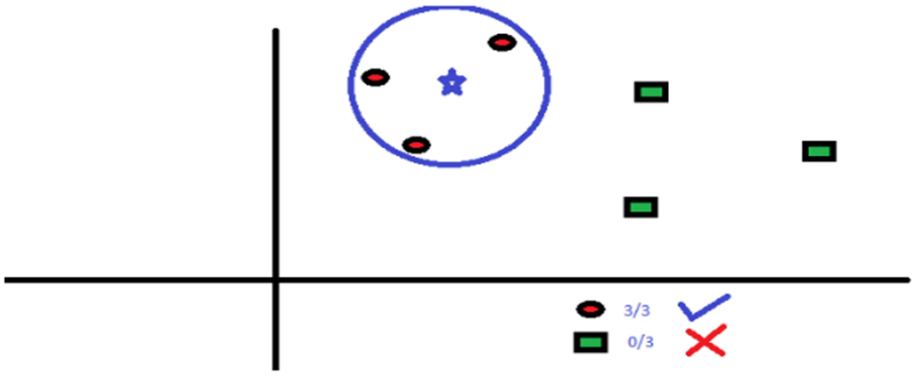

Figure 2.4 Classification Using KNN [21]

2.3.2 Linear Regression (LR)

LR efforts to system the association between two parameters by fitting a linear model to experiential facts. One parameter is measured to be an instructive parameter, and the other is measured to be a reliant on parameter. For instance, a modeler might poverty to narrate the weights of entities to their statures using a linear regression equation [8].

Before trying to fit a linear equation to experiential data, a modeler must first define whether or not there is an association between the parameters of attention. This does not unavoidably suggest that one parameter, but that there is nearly important connotation between the 2 parameters. A scatterplot can be an obliging instrument in decisive the forte of the association between 2 parameters. If there seems to be no connotation between the future descriptive and reliant on parameters, then fitting a linear model to the data perhaps will not deliver a beneficial system. A valued arithmetical amount of connotation between 2 parameters is the correlation constant, which is a rate between -1 and 1 representative the forte of the connotation of the experiential information for the 2 parameters.

A linear model line has a system of the formula:

$$f(X) = a + mX \tag{2.1}$$

where X is the descriptive mutable and f(X) is the reliant on parameter. The slope of the line is m, and a is the interrupt. The linear regression curve shown in the Figure 2.5.

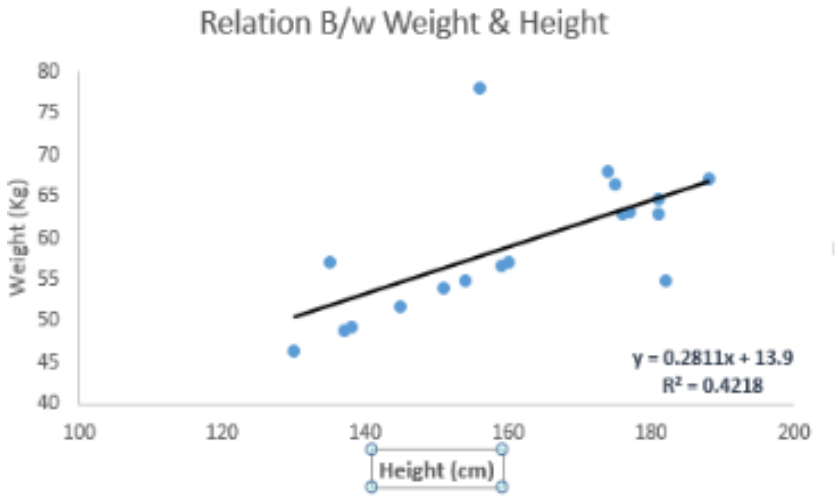

Figure 2.5 Linear Regression Model [23]

Furthermore, Logistic regression is applied to discover the possibility of (event=Success) and (event=Failure). We must apply logistic regression when the reliant on parameter is binary (0/1) in countryside. Now the rate of Y varieties between 0 to 1 and it can have signified by Figure 2.6.

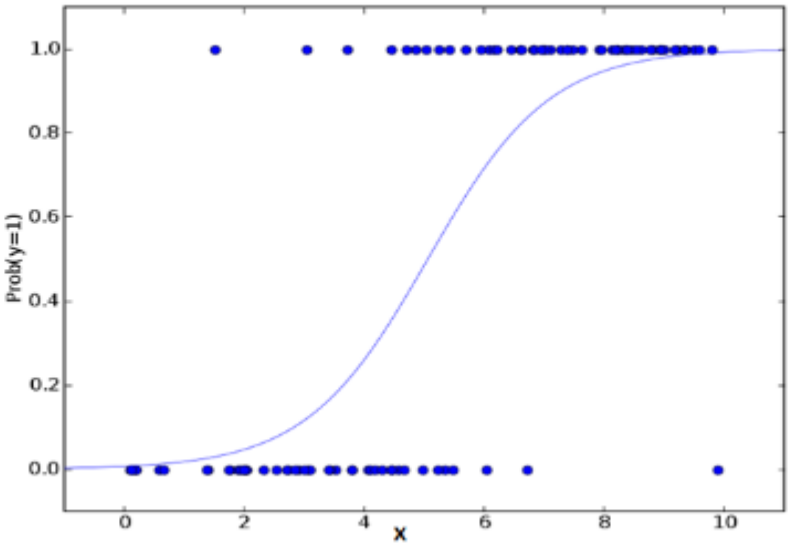

Figure 2.6 Non-Linear Regression Model

Moreover, a regression equation is a polynomial regression model if the control of self-governing parameter is greater than 1. The Figure 2.7 shows a polynomial equation:

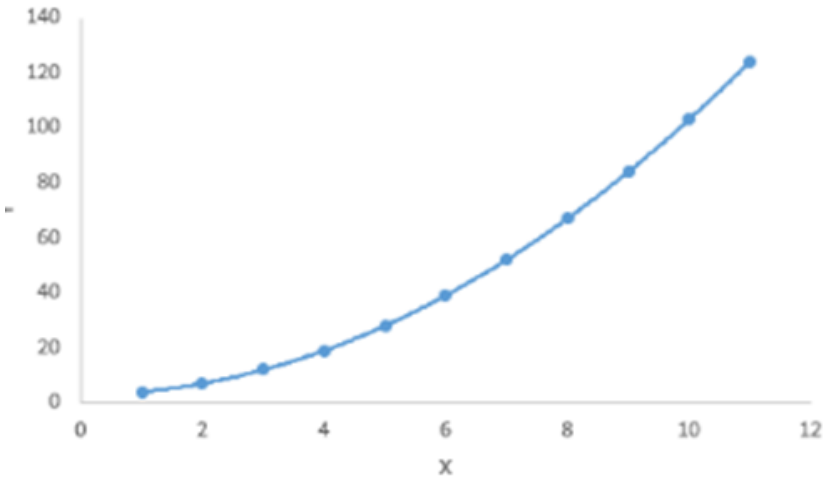

Figure 2.7 Polynomial Regression Model

2.3.3 Support Vector Machine (SVM)

SVM is a discriminative classifier officially definite by a splitting hyperplane. In our study, assumed labeled training features, the technique productions a best hyperplane which classifies new instances. In two dimensional interplanetary this hyperplane is a line separating a plane in 2 sections where in each lesson lay in either side. Then, classifying any data have more than one Possible hyperplane. In Figure 2.8 and 2.9 the possible hyperplanes presented for one equation [9].

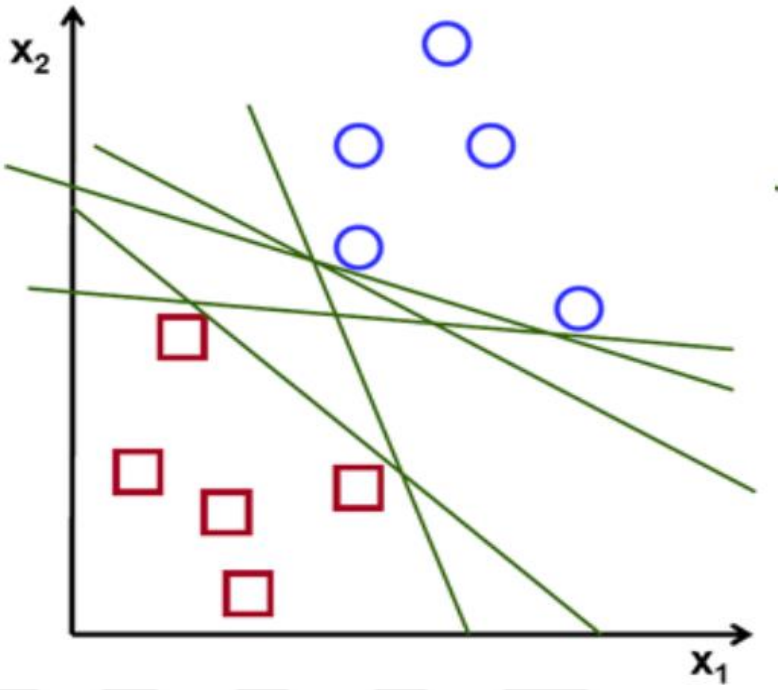

Figure 2.8: Possible Hyperplane 1

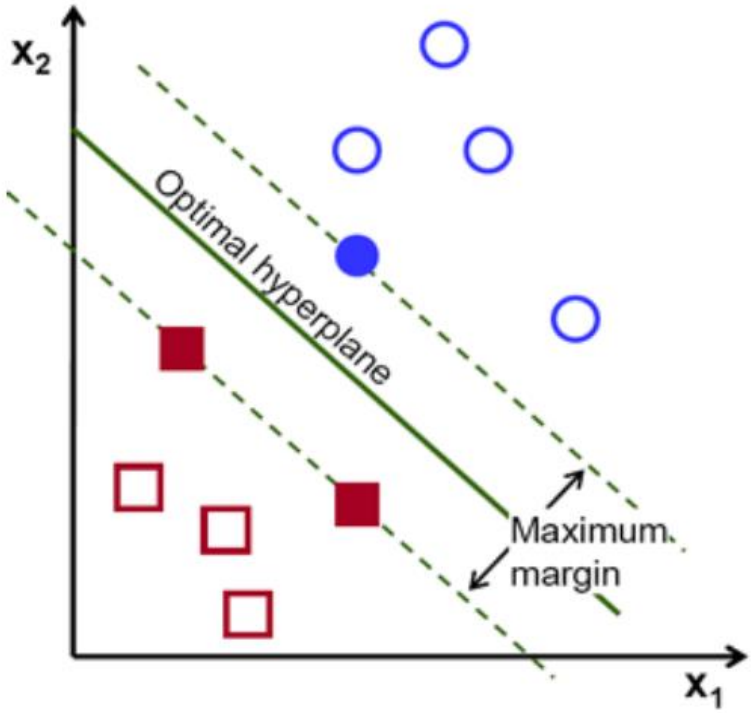

Figure 2.9: Possible Hyperplane 2

Using these SVM, we increase the margin of the classifier. Deleting the support vectors will modification the location of the hyperplane. The margin is very important issue when building SVM classifier.

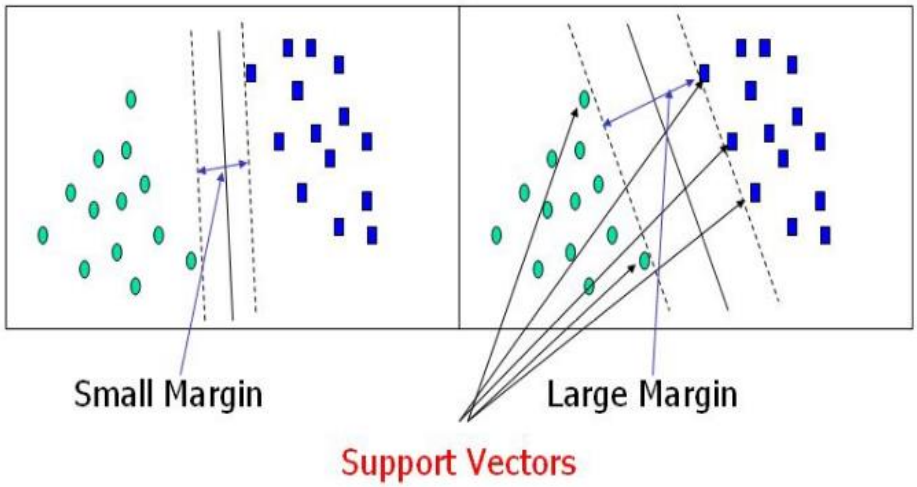

Figure 2.10 SVM Margin Size

SVM classifies a binary problem applying a linear hyperplane by presumptuous that the training set has n-training examples, that is, (x_1 , y_1), (x_2 , y_2), ..., (x_n , y_n), where x i ∈ N is an N dimensional vector that fits to one of lessons y i ∈ {−1 , +1 }. The definite binary cataloging problematic can be detached applying a linear decision model presented in Eq (2.2):

$$f(x) = w.x + b \tag{2.2}$$

Which w represented vector the selected the orientation of the desired hyperplane, b represented the bias. The best hyperplane for separate two objects represented as shown in Eq (2.3)

$$y_i(w.x + b) \geq 1 \tag{2.3}$$

The result to this problematic can be found by resolving the following forced optimization problem, and the aim of SVM process is to reduce the equation that presented in Eq (2.4):

$$\text{Minimize } \frac{1}{2} w.w + c \sum_{i=1}^{n} \varepsilon_i \tag{2.4}$$

2.3.4 Decision Tree (DT)

A DT has several equivalences in actual life, and tries out that it has prejudiced an extensive part of machine learning, cover mutually classification and regression. In decision study, a decision tree can be applied to visually and openly signify decisions and decision creation. As the term energies, it applies a tree-such as model of decisions. However, a usually applied tool in data mining for originating a plan to spread a specific objective, it's also extensively applied in machine learning [10].

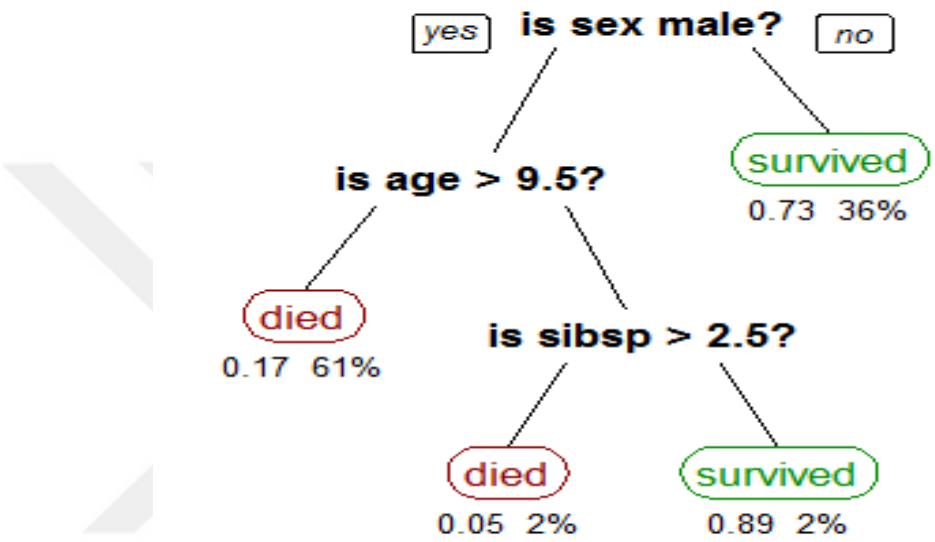

Figure 2.11 Decision Dree [26]

2.3.5 Factor Analysis

D denotes the data set which contains of D-dimension actual standards D= {y_1, y_2, y_3}. The input dataset is formed by applying the Equation (2.5).

$$y= \Lambda x+ \varepsilon \tag{2.5}$$

where x is a k-dimensional zero-mean Gaussian vector with data competition to concealed features, Λ is atmosphere of factors in the method of D X K. The particulars of the mathematical model were illustrated in [11].

2.3.6 K-Means Clustering

K-Means clustering aims to divider n items into k groups in which individually item belongs to the cluster with the adjacent mean. This technique yields exactly k dissimilar clusters of highest conceivable difference. The greatest amount of clusters k foremost to the utmost parting (space) is not identified as a priori and necessity be calculated from the feature. The purpose of K-Means clustering is to reduce whole intra-cluster alteration. The procedures of the technique presented below [30].

1. Clusters the features into k sets where k is determined by the user.
2. determine k facts at accidental as cluster middles.
3. Assign objects to their closest cluster center according to the *Euclidean distance* function.
4. Compute the centroid or mean of all items in every cluster.
5. Recurrence stages 2, 3 and 4 till the similar facts are allocated to every cluster in successive circles.

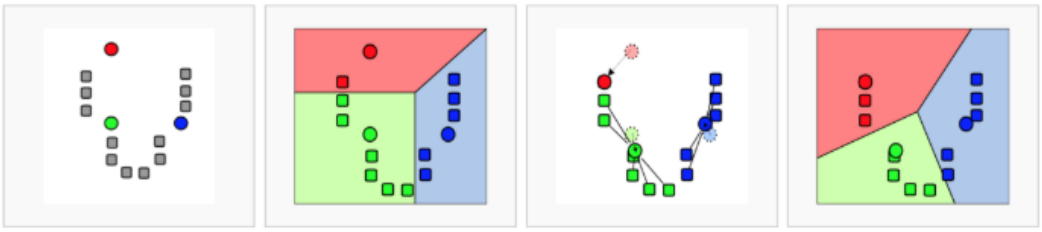

Figure 2.12 K-means Clustering

K-Means is comparatively a well-organized technique. Though, we essential to stipulate the numeral of clusters, in spread and the last outcomes are subtle to initialization and often dismisses at a local optimum. Inappropriately, there is no general theoretical technique to discover the best amount of clusters. An applied method is to associate the consequences of manifold runs with dissimilar k and select the optimal one depends on a predefined standard. In overall, a great k perhaps reductions the error but upsurges the danger of overfitting [12].

2.3.7 Principal Component Analysis (PCA)

PCA is a common technique applied in statistical learning methods. PCA can be applied to realize dimensionality decrease in regression settings letting us to explain a high-dimensional data set with a minor numeral of illustrative features which, in mixture, designate greatest of the erraticism originate in the original high-dimensional data.

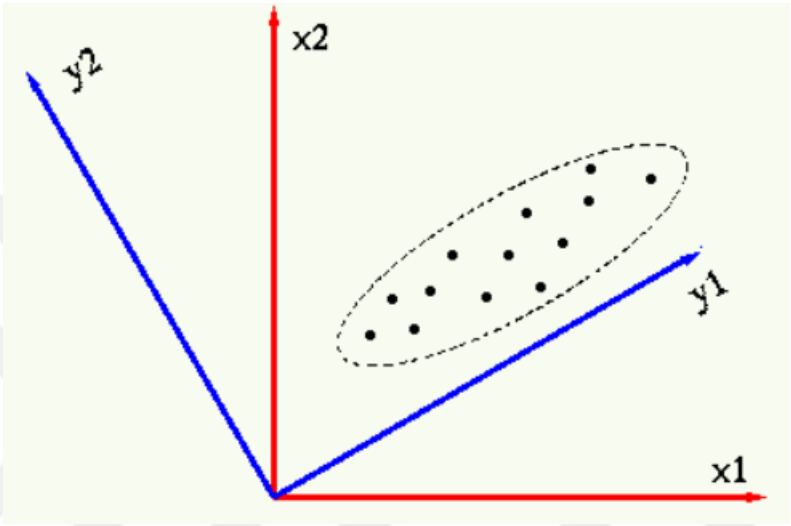

Figure 2.13 Principal Component Analysis

PCA can besides be applied in unsupervised learning fields to learn, imagine a discover patterns in high-dimensional data sets when there is not exact reply feature. Moreover, PCA can assistance in clustering trainings and subdivision systems [13]. The method of PCA presented in Figure 2.13.

2.4 Litreature Review

In this section, several previous studies are presented and explained, all these studies applied machine learning and deep learning techniques to deal with in gastric cancer. Toshiaki Hirasawa et al. presented CNN based on Single Shot MultiBox structure which trained using 13,584 gastric cancer endoscopic images.

The proposed system validated using 2296 gastric cancer endoscopic images which collected from 69 cases To evaluate the diagnostic accuracy, an independent test set

of 2296 whith 77 gastric cancer gashes was used to the built CNN. The proposed CNN model complete the testing for 2296 within 47s which mean its very fast and presented 98.6% accuracy [14].

Furthermore, Zhang X et al. proposed CNN based Single Shot MultiBox Detector (SSD) structure too for Gastric Polyps recognition. The SSD based CNN is real time polyp detection system which can analysis 50 frame in second and increase the detection precision to 90.4% from 88.5% [15].

Moreover, Y. Sakai et al. proposed CNN for automatic gastric cancer detection using endoscopic images. The transfer learning assist the CNN and presented 87.6% which is satisfactory result when compared with our studies [16].

Tourdot et al.[33] presented that the analog 1Y of HER-2650-658 has sturdier empathy to HLA-A2 than WT epitope. previ

Furthermore, Andersenet al. [34] established that adapting epitope orders by presenting a methionine in the additional location of Survivin 96-104 canenhance peptide immunogenicity. The competence of this technique wasverified by changing analog 2 M of the TRP-2180-188 peptide, which canenhance peptide compulsory and sluggish its detachment after HLA-A2[35].

Previous, Zaremba et al. presented that analog CAP1-6D, which appears tobelong to a dissimilar period of relieved cytotoxic T lymphocyte (CTL)peptides, can improve TCR credit of the peptide-MHC complexwithout cumulative MHC compulsory, and can consequently be applied to change thevaccine epitope[36].

In [37] Garcia et al. proposed a method based on CNN for Automatic lymphocyte recognition on IHC pictures of gastric cancer. The database produced as portion of this study is overtly obtainable for future study.

In [38] Cao et al. applied the Mask R-CNN technique to recognize the compulsive units of gastric cancer, and section the cancer shell, and then enhance it by regulating factors. The technique lastly lets it to get a validation outcome with an AP worth of 61.2 when recognizing medical images.

CHAPTER 3

MATERIAL AND METHODS

In this chapter, several deep learning techniques are presented and explained in detail form. Furthermore, the proposed method is explained with flowchart.

3.1 Deep Learning (DL)

Deep learning is version of a larger family of machine learning techniques depended on artificial neural network (ANN). DL have several techniques such as DNN, deep belief networks (DBN), recurrent neural networks (RNN) and convolutional neural networks (CNN) have been used to area containing computer vision, speech recognition, bioinformatics, medical image recognition, where they presented outcome comparable to and in some cases higher to human experts [17]. Number of deep learning techniques is presented below:

3.1.1 Convolutional Neural Network (CNN)

CNN mimics the human eyes system so its presented high and remarkable results in computer vision applications. The CNN consist from several operation that lead to extracted high level features by using shared weight and learning local features from images, video and audio data etc.

As shown in Figure 3.1 input layer mean the data that will be analyzed by the network, convolutional layer is used to extracted features and used share Wight technique for optimal using the memory. Pooling layer used to reduce the dimension of the data. The last stage is the fully connected layer which is used to classify the linear extracted features in to labels [18].

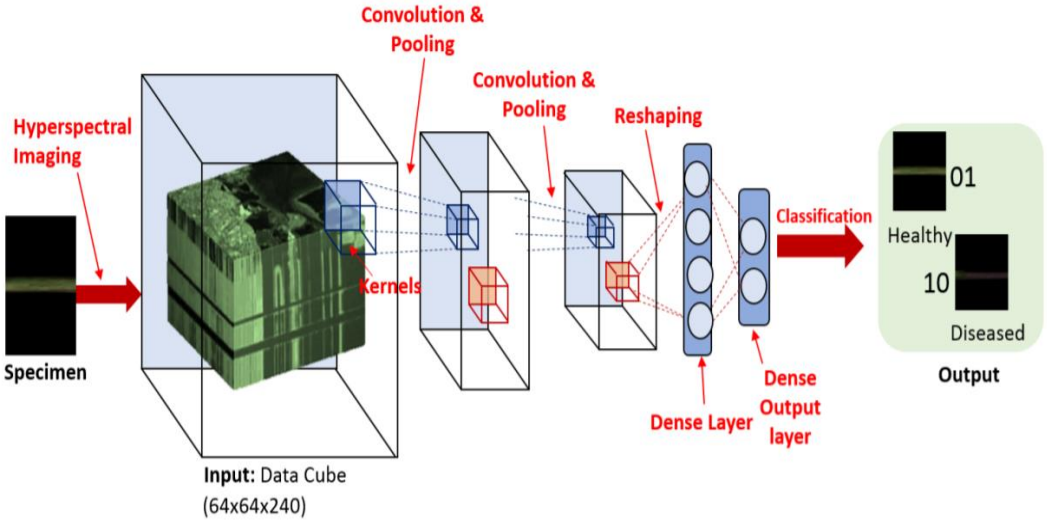

Figure 3.1 Convolutional Neural Network [18]

- **Convolutional Layer**

The purpose of a Convolutional layer is to mine and learn subtle features of the input data. Individual a helping of the effort data is linked to the following Convolutional layer since the usage of all information principal to too abundant calculation time. Then, it smears dot crops between designated variety on information and a sieve on all the sizes [19]. The crop of this process is the feature chart. The sieve transparencies over the input image using stride function see Figure 3.2.

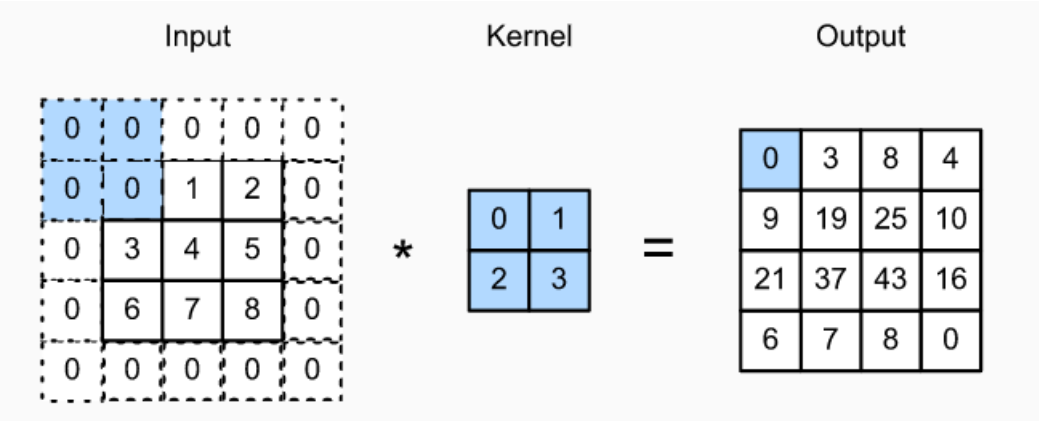

Figure 3.2 Convolution Layer

- **Pooling**

The pooling layer is introduced among convolution layer in CNN architecture. The pooling is a function to decrease the longitudinal size of the features to decrease the amount of parameters and calculations in NN then control the overfitting.

The greatest mutual form is a pooling layer with filters of size 2x2 or any size depend on the input belongings used with a stride integer number 2,3,4 down examples all complexity helping in the features by any number lengthways both width and height, elimination 75% of the activations [20].

Usually, two types of pooling are applying the max pooling and average pooling approaches. The max pooling funds choosing the upper value data in each stride. In this instance 2*2 filter with 2 strides are deliberate see Figure 3.3.

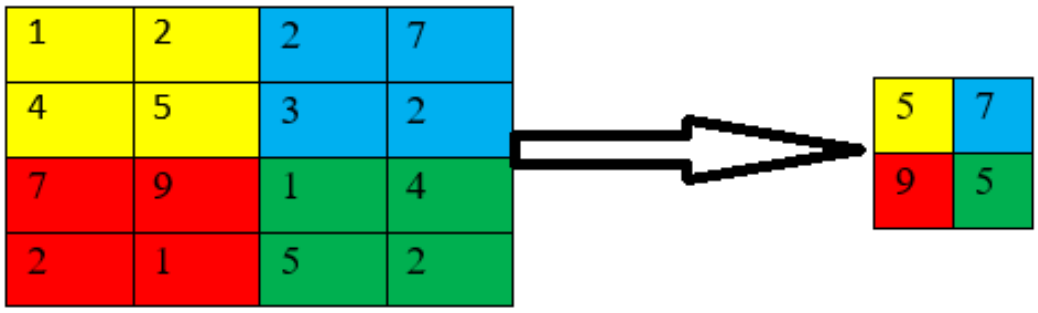

Figure 3.3 Max Pooling

Furthermore, the similar instance applied with average pooling and obtainable other results see Figure 3.4.

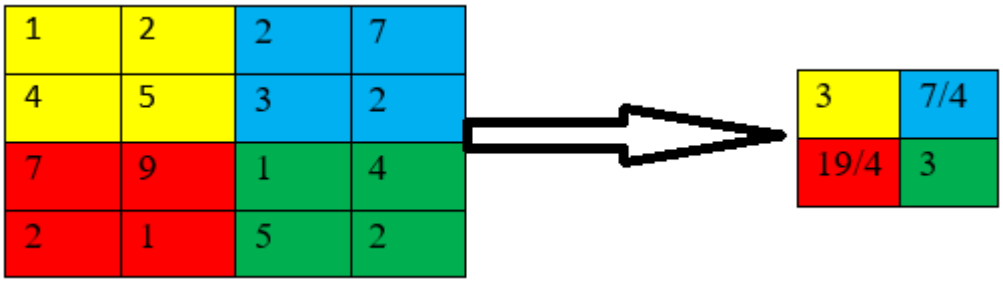

Figure 3.4: Average Pooling

In each stride the data are composed and alienated on the amount of features, the applying of max pooling or average pooling depended on the feature kind apiece of them obtainable acceptable results rendering to the topographies conduct.

- **Fully connected layer (FC)**

The FC layer is like a customary NN but in deep learning its applied in the last layer after selected the important features and amount of feature input are decreased. The FC layer mean connect each node in dissimilar layers with each other. Commonly, SoftMax applied as transfer function in this section see Figure 3.5.

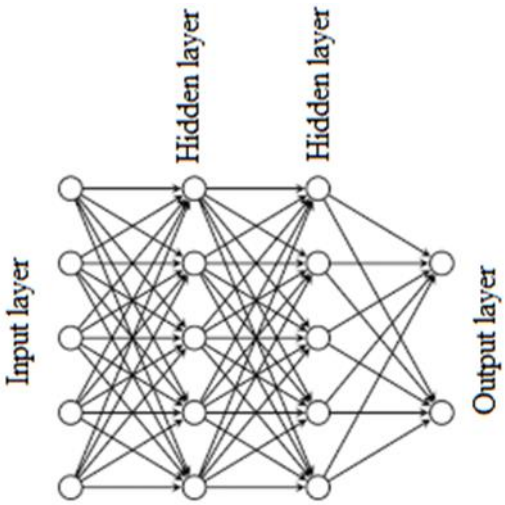

Figure 3.5 Fully Connected Layer

Normally, FC layer contains of three layers as input, hidden and output layers and there is a capability to upsurge the amount of hidden layer rendering to the data performance and size [21].

3.1.2 Recurrent Neural Network (RNN)

The impression behindhand RNNs is to brand apply of sequential info. In an outdated neural network, we shoulder that all contributions (and outputs) are sovereign of apiece additional. Nonetheless for numerous errands that's an actual wicked impression. If you poverty to expect the following term in a sentence you healthier distinguish which arguments came beforehand it. RNNs are named recurrent since they achieve the

similar task for each component of a sequence, with the production being been depending on the earlier calculations. Additional method to think about RNNs is that they have a "memory" which detentions data about what has remained intended so remote [22]. In theory RNNs can brand use of data in randomly extended orders, but in repetition they are incomplete to observing back lone insufficient ladders. See Figure 3.6.

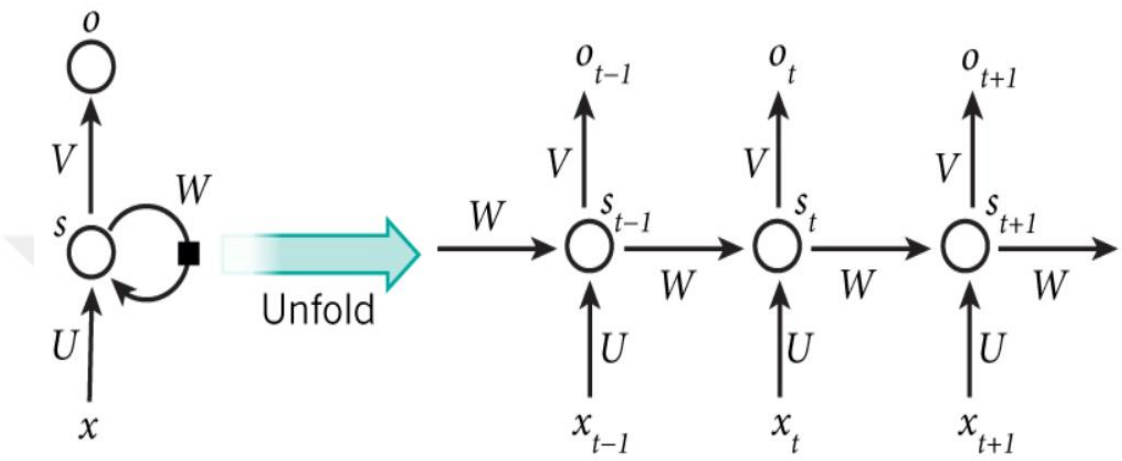

Figure 3.6 RNN Structure

The Figure 3.4 shows RNN existence opened (or unfolded) hooked on a full system. By opening we just mean that we inscribe out the net for the whole sequence. For instance, if the sequence we care around is a sentence of 5 words, the system would be opened hooked on a 5-layer NN, one layer for separately word [23]. There are two advantages of using RNN:

- An RNN recalls apiece and each data finished time. It is valuable in time sequence forecast only since of the feature to recall preceding inputs as healthy.
- RNN are smooth castoff with convolutional layers to spread the real pixel neighbourhood.

3.2 Long Short Term Memory Networks (LSTM)

To resolve the tricky of Disappearing and Explosion Gradients in a deep RNN, numerous differences were industrialized. Unique of the greatest well-known of them

is the LSTM. In idea, an LSTM recurrent component stabs to “remember” all the historical information that the network is understood so distant and to “forget” immaterial facts. This is complete by presenting dissimilar activation function layers named “gates” for dissimilar drives. Separately LSTM recurrent component also upholds a vector named the Interior Cell State which theoretically labels the data that was selected to be booked by the preceding LSTM recurrent element [24]. See Figure 3.7.

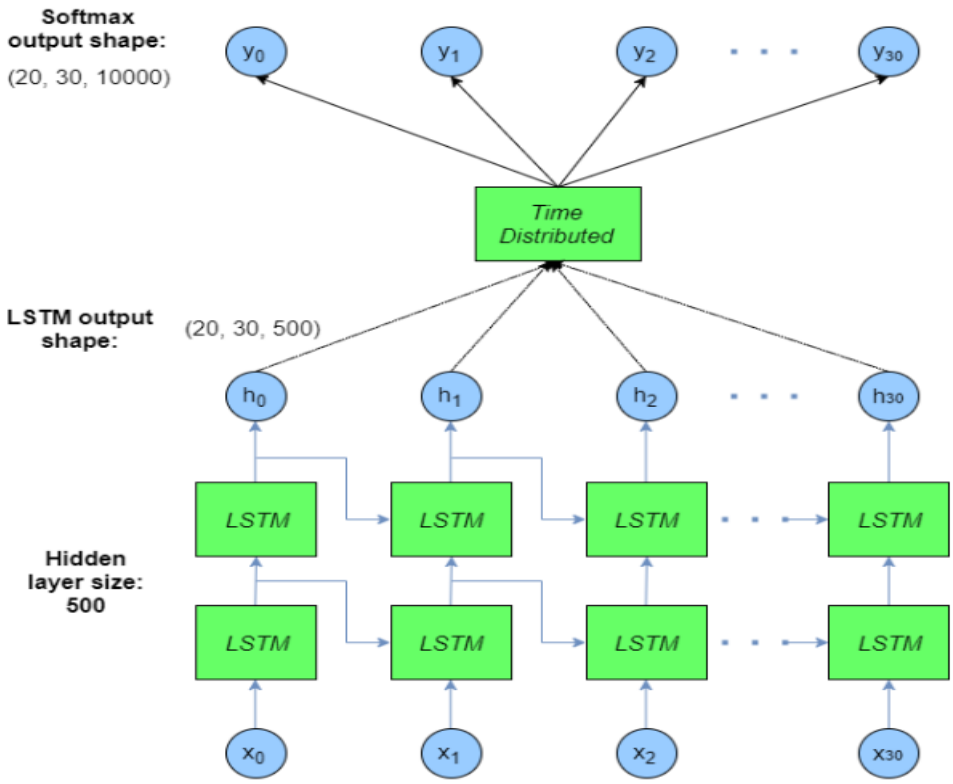

Figure 3.7 LSTM Based SoftMax

3.3 Transfer Learning

Transfer learning is an essential instrument in ML to solve the fundamental problem of unsatisfactory training data. It tries to transfer the knowledge from the supply field to the target field by relaxing the belief that the training and testing data, which will

lead to a huge positive impact on several fields that are hard to enhance because of unsatisfactory training data [25].

There are several pre-trained organizations are offered, the reasons for using pre-trained models are presented below:

- First, it needs additional computational power to train the large models on huge datasets.
- Second, high time taking to train the network up to several of days.

There are several categories of transfer learning listed below:

- Instances-based deep transfer learning:
  Instances-based deep transfer learning means to apply a particular weight correction approach, choose limited examples from the resource area as complements to the training group in the objective area by transferring suitable weight principles to these select examples.

- Mapping-based deep transfer learning: Mapping-based deep transfer learning describes to mapping examples after the source field and objective field into a new features area. In this new data area, examples from two areas are likely and appropriate for a society DNN.

- Network-based deep transfer learning: Network-based deep transfer learning indicates to the reprocess the limited network that pre-trained in the supplier area, involving its network organization and relation parameters, transfer it to be a part of DNN which applied in target area. It pretends that" NN is comparable to the handling mechanism of the human brain, and it is an iterative and uninterrupted generalization procedure.

Several pre-trained networks are presented by researcher and these networks are applied in many classification problems [26].

3.3.1 AlexNet

AlexNet is the name of a CNN, presented by Alex Krizhevsky, AlexNet participated in the ImageNet huge Scale Visual Recognition Task on 2012. The network attained a top-5 error of 15.3%, further than 10.8% goals lower than that of the execution.

AlexNet included eight layers; were convolutional layers represented the first five, several of them followed by max-pooling layers, fc layers represented in the last three were.[27] It applied the non- ReLU activation function, which indicated enhanced training accuracy over tanh and sigmoid. The AlexNet structure presented in the Figure 3.8.

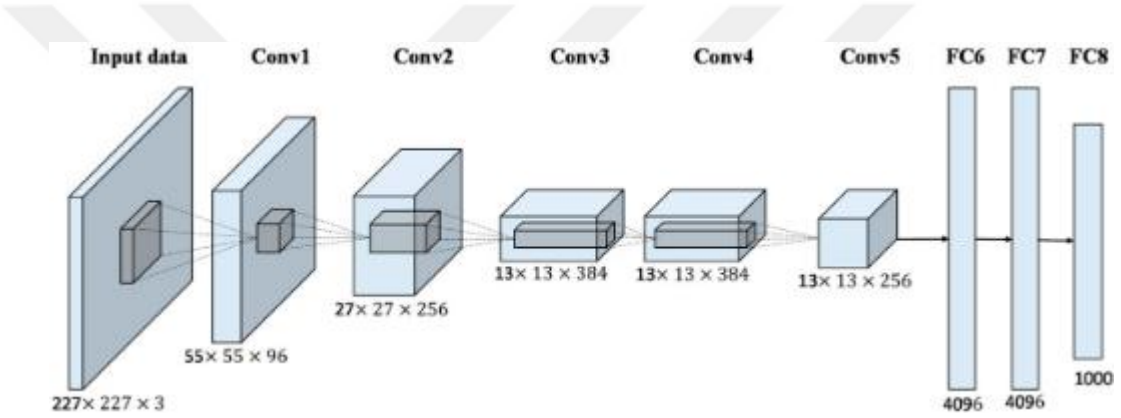

Figure 3.8 AlexNet Structure

3.3.2 GoogleNet

The major improvement of GoogleNet is apply of a structure termed Inception [29]. In common, Inception is a net in net architecture, and the optimum local sparse architecture of a vision net is spatially recurring from the beginning to the ending. Three Inception architecture applied in various environments are presented: typically, 1 × 1 convolution is applied in Inception to calculate cuts before the expensive 3 × 3 and 5 × 5 convolutions [28]. The structure of GoogleNet represented in Figure 3.9.

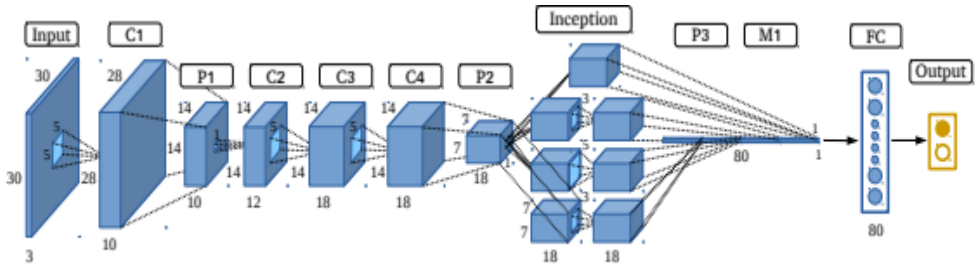

Figure 3.9 GoogleNet Structure

3.3.3 ResNet101

So far, while improving the net depth necessarily accuracy also improves. But various problems occur near with net depth in ResNet. Improved depth that necessitated altering the weights, which occurs the ending of the network, the estimate enhances small at the preliminary layers [29]. An Extra one is large parameter area it needed. The structure of ResNet101 presented in Figure 3.10.

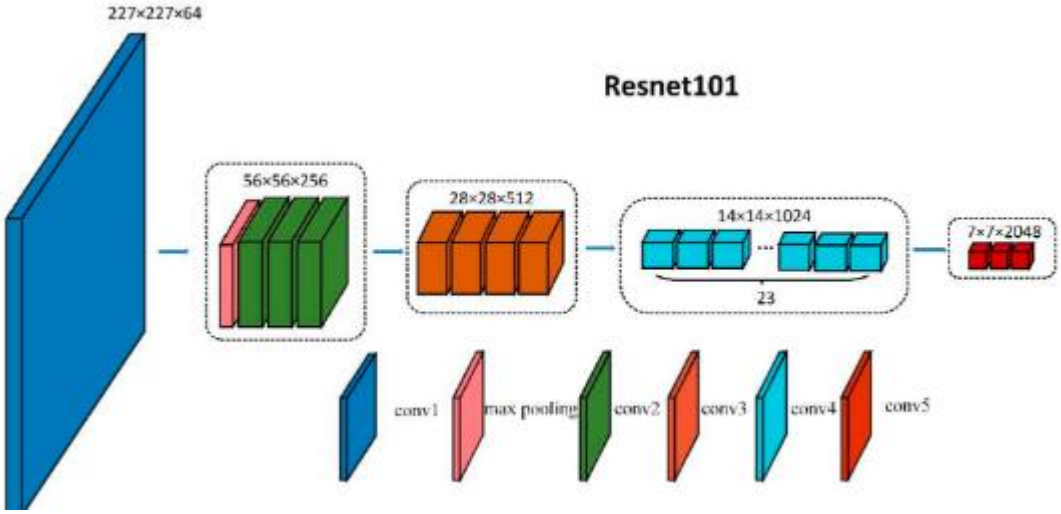

Figure 3.10 ResNet101 Structure

3.3.4 DenseNet201

DenseNet was recognized for the astonishing presentation on the cutthroat object recognition standard tasks like ResNet101 and AlexNet. The initial effort to solve huge scale image recognition task by deep CNN, earned the first place and second place in

localization and recognition assignment separately. DenseNets have numerous exciting benefits: they improve the vanishing-gradient problem, improve feature reproduction, promote feature recycle, and significantly decrease the number of parameters [30]. The structure of DenseNets201 shown in Figure 3.11.

Layers	Output Size	DenseNet-121($k = 32$)	DenseNet-169($k = 32$)	DenseNet-201($k = 32$)	DenseNet-161($k = 48$)
Convolution	112 × 112	7 × 7 conv, stride 2			
Pooling	56 × 56	3 × 3 max pool, stride 2			
Dense Block (1)	56 × 56	[1 × 1 conv; 3 × 3 conv] × 6	[1 × 1 conv; 3 × 3 conv] × 6	[1 × 1 conv; 3 × 3 conv] × 6	[1 × 1 conv; 3 × 3 conv] × 6
Transition Layer (1)	56 × 56	1 × 1 conv			
	28 × 28	2 × 2 average pool, stride 2			
Dense Block (2)	28 × 28	[1 × 1 conv; 3 × 3 conv] × 12	[1 × 1 conv; 3 × 3 conv] × 12	[1 × 1 conv; 3 × 3 conv] × 12	[1 × 1 conv; 3 × 3 conv] × 12
Transition Layer (2)	28 × 28	1 × 1 conv			
	14 × 14	2 × 2 average pool, stride 2			
Dense Block (3)	14 × 14	[1 × 1 conv; 3 × 3 conv] × 24	[1 × 1 conv; 3 × 3 conv] × 32	[1 × 1 conv; 3 × 3 conv] × 48	[1 × 1 conv; 3 × 3 conv] × 36
Transition Layer (3)	14 × 14	1 × 1 conv			
	7 × 7	2 × 2 average pool, stride 2			
Dense Block (4)	7 × 7	[1 × 1 conv; 3 × 3 conv] × 16	[1 × 1 conv; 3 × 3 conv] × 32	[1 × 1 conv; 3 × 3 conv] × 32	[1 × 1 conv; 3 × 3 conv] × 24
Classification Layer	1 × 1	7 × 7 global average pool			
		1000D fully-connected, softmax			

Figure 3.11 DenseNet201 Structure

3.3.5 VGG-16

Simonyan and Zisserman of the University of Oxford designed a 16-layer where 16 convolutional layers, three fully-connected. CNN that rigorously used stride with 3×3 filters and pad of 1, max-pooling with 2×2 filter with stride 2, called VGG-16. To decrease the total of parameters in CNN, it applies 3×3 small filters in whole convolutional layers and top used with its 7.3% error rate [31]. The VGG-19 model applied in various image recognition, video classification and object detection etc. The structure of VGG-16 shown in Figure 3.12.

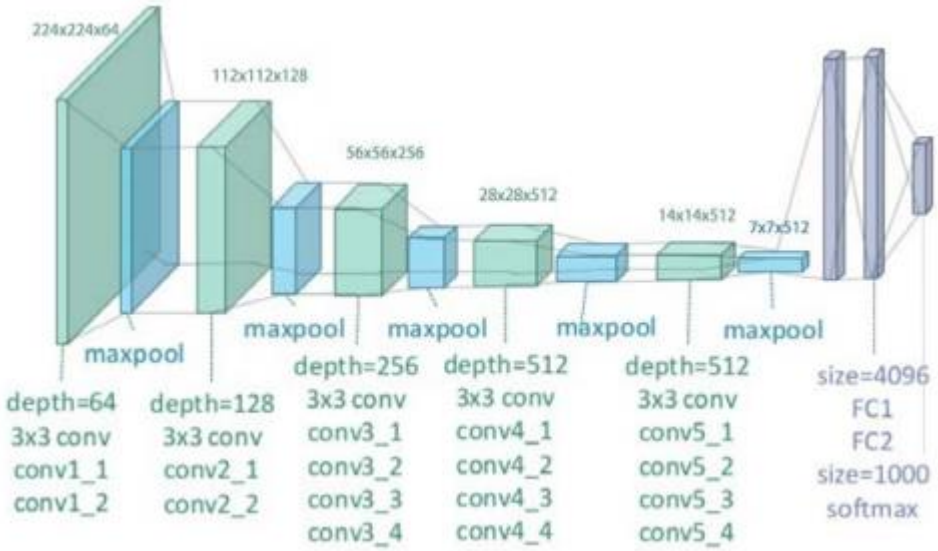

Figure 3.12 VGG16 Structure

3.3.6 OurNet

In this thesis, we presented efficient network trained and tested for Gastric Cancer detection. The presented CNN structure consists from 8 layers which consist from input layer, convolution layers, pooling layers, and fully connected layer. As shown in the Figure 3.13 the input layer consists from 224*224*3 which each input image normalized automatically to this standard size. Then, three convolution and three pooling layers are followed after input layer. In the last layer fully-connected layer used to represent high level linear features. The size of fully connected layers equal to the number of classes.

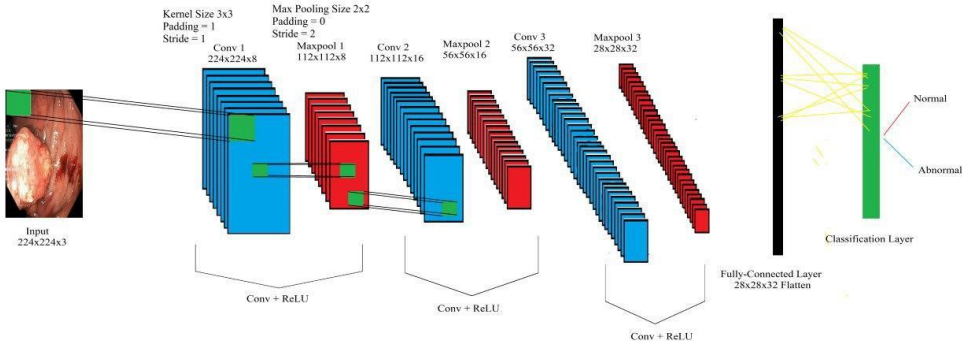

Figure 3.13 OurNet Structure

Moreover, the structure of the OurNet presented in detail form in the Table 4.1. The size of the filter and output size of each layer presented too. The filter size is the important factor which affected the accuracy of the network and the output size of each layer.

Table 4.1 OurNet Structure

Type	Filter size / Stride	Output size
Input image		224x224x3
Conv_1	3x3 / 1	224x224x8
Maxpool_1	2x2 / 2	112x112x8
Conv_2	3x3 / 1	112x112x16
Maxpool_2	2x2 / 2	56x56x16
Conv_3	3x3 / 1	56x56x32
Maxpool_3	2x2 / 2	28x28x32
Fully_Connected Layer		1x1x2
Softmax		1x1x2

3.4 Dataset

In this study, the dataset created by collecting two common used datasets contains Gastric Cancer images which abnormal data obtained from [39] and all dataset available in https://datasets.simula.no/kvasir/ . The Kvasir dataset contains of pictures, explained and proved by doctors (expert endoscopists), counting numerous classes display anatomical landmarks, pathological discoveries or endoscopic processes in the GI tract, i.e., tens of images for individually class. The amount of images is enough to be applied for various tasks such as data mining, machine transfer learning, and deep learning. Furthermore, the dataset involves of the images with dissimilar sizes between 1920x1072 and 720x576 pixels. The original dataset consists from 1000 images but in this study only 750 of these images are selected.

On the other hand, normal data which represented the data without cancer problem obtained from [15] and the full dataset can be found at

https://github.com/jiquan/Dataset-acess-for-PLOS-ONE. Furthermore, the original dataset consists from 758 images which only 750 of them are selected in this study. Then, our dataset consists from 1500 images with two classes which each class consist from 750 images. All images sizes are normalized and the sizes are fixed to 224*224. The aim of resize all image and fixed the dimensions in one size is to become easy to used when applied machine learning and deep learning techniques to deal with this dataset.

3.1.5 Proposed Method

In this study, the Gastric cancer detection system presented by combining a CNN based SoftMax classifier in the last layer. The CNN applied to extract and learn the input image features; CNN is applied, taking benefit of the ability to share weight idea to learn more sensitive feature with low number of parameters. So, in this thesis, AlexNet, GoogleNet, ResNet101, Vgg16 and DenseNet201 was adapted and applied for feature extraction and a multi-class SoftMax was used for recognition on a single model which can be seen in Figure. 3.13.

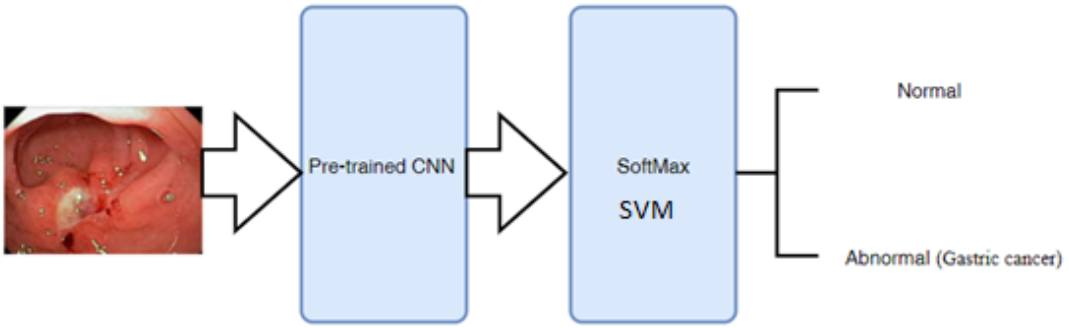

Figure 3.14 Block Diagram of Proposed Method

SoftMax function computes the probabilities distribution of the occurrence above ‘n’ various occurrences. In overall way of saying, this function will compute the probabilities of individually label class over all possible label classes. Then, the computed probabilities will be useful for selecting the label class for each input.

The core advantage of using SoftMax is the output probabilities domain. The domain is 0 to 1, and the total rate of the probabilities will become equal to one. When the

SoftMax function used as multi-classifier model it produces the probabilities of each label and the correct label will have the high probability.

The mathematic model of SoftMax function is presented below, where the vector of inputs to the output layer represented by z. And again, j indexes the output units, so i = 1, 2, ..., K see equation 3.1.

$$SoftMax(y) = \frac{e^{zi}}{\sum_{k=1}^{k} e^{zak}} \tag{3.1}$$

The SVM mathematically represented as shown in equations 4.1-4.6.

$$\overrightarrow{x} = (x_1, \dots., x_p) \in R^P \tag{3.2}$$

The equation 4.1 represented the elements of p-dimensional feature and the hyperplane represented in the equation 4.2.

$$b_0 + b_1 x_1 + \cdots + b_p x_p = 0 \tag{3.3}$$

$b_0 \neq 0$ Gives us an affine plane. We can use a more concise notation for this equation by introducing the summation sign:

$$b_0 + \sum_{i=1}^{p} b_i \, x_i = 0 \tag{3.4}$$

$$\vec{b}.\vec{x} + b_0 = 0 \tag{3.5}$$

Elements $\vec{x}$ above the plane satisfy:

$$\vec{b}.\vec{x} + b_0 > 0 \tag{3.6}$$

While those below it satisfies:

$$\vec{b}.\vec{x} + b_0 < 0 \tag{3.7}$$

The flowchart of the proposed method presented in Figure 3.15.

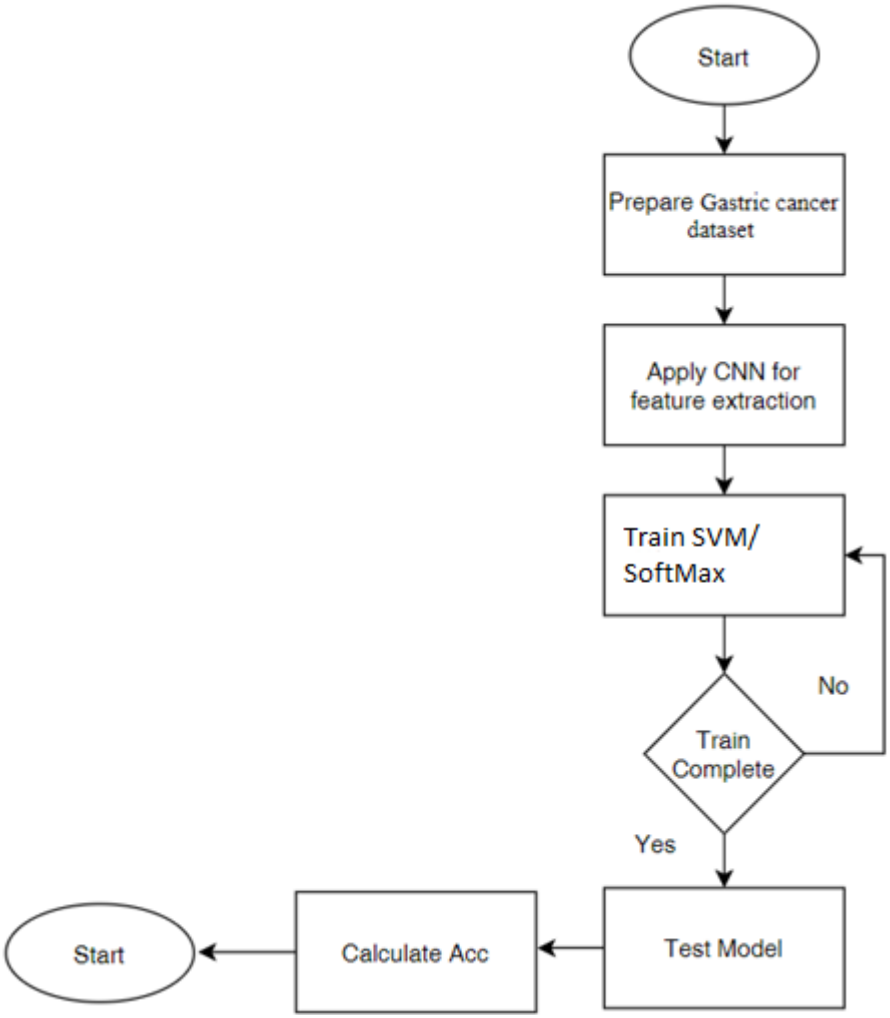

Figure 3.15 Flowchart of Proposed Method

CHAPTER 4

RESULTS

In this chapter, the experiment will be presented along with the result and the evaluation, divided into sections as the follows:

4.1 MATLAB Tool

MATLAB is a user interface design tool intended precisely for engineers and researchers? The feeling of MATLAB is the MATLAB language, a matrix-based programing language permitting the highest usual arrival of computational arithmetic.

- Data analyze
- Algorithms develop
- Create applications and models

The language, apps, and built-in math functions allow the researcher to speedily discover manifold methods to reach at a solution. MATLAB lets the experts yield your thoughts from investigation to manufacture by positioning to initiative applications and entrenched strategies, as well as mixing with Simulink® and Model Based Design.

4.2 IMPLEMENTATION

4.2.1 Hold-Out

Hold-out is once the developer divided the dataset into a ‘train’ and ‘test’ groups. The training group is what the method is trained on, and the test group is applied to get how well that method achieves on unobserved facts. A mutual divided when applying the hold-out technique is applying 20% of the data for testing and 80% of data for training the method. See figure 4.1 for showing the Hold-out procedure.

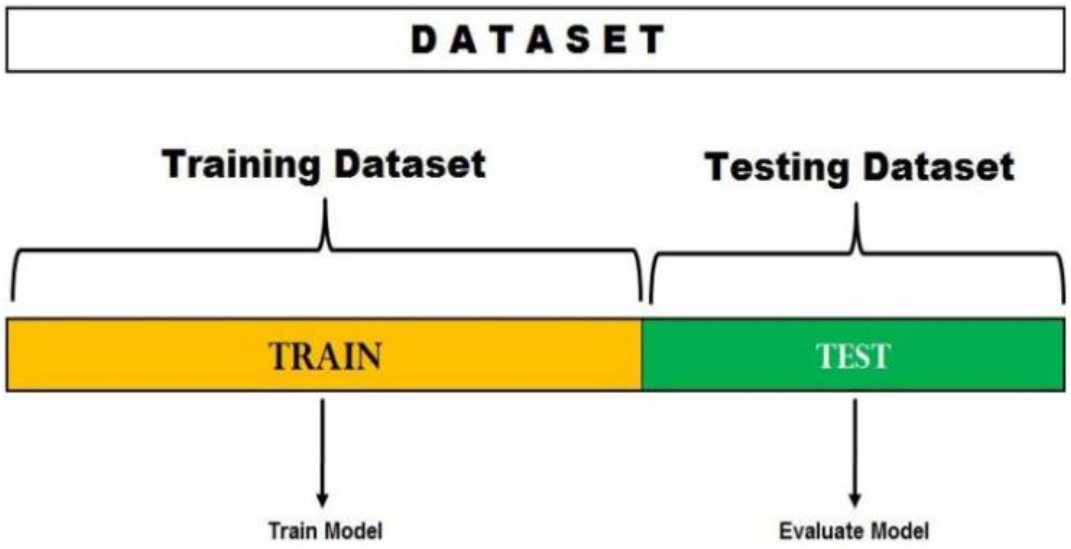

Figure 4.1 Hold-out

4.2.2 K-Fold Cross

K-Fold mean specified data group is divided into a K integer of folds wherever individually fold is applied as a testing group at about opinion. For example, take scenario of 5-Fold cross validation (K=5). Now, the data group is divided into 5 sections. In the first epoch, the first section is applied to validate the method and the remain are applied to train the method. In the second epoch, 2nd section is applied as the testing group while the remain serve as the training group. This procedure is recurrent till for each section of the 5 section have been applied as the testing group.

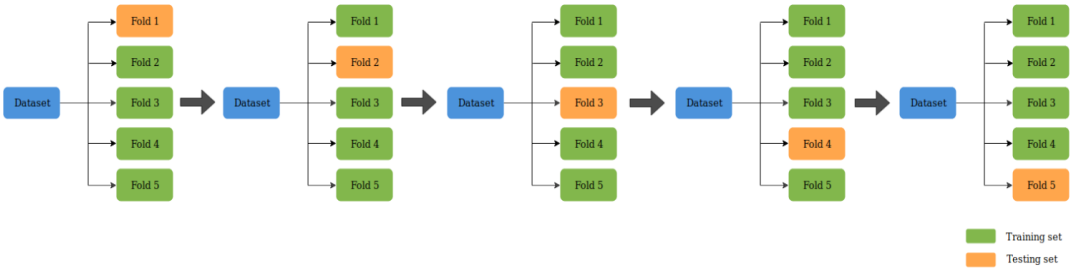

Figure 4.2 K-fold Cross

4.2.3 Random Subsampling

In this method, several groups of data are randomly selected from the database and mixed to form a test database. The rest data used to train the method. The following diagram denotes the random subsampling validation method. The accuracy of the method is the average of the accuracy of each epoch.

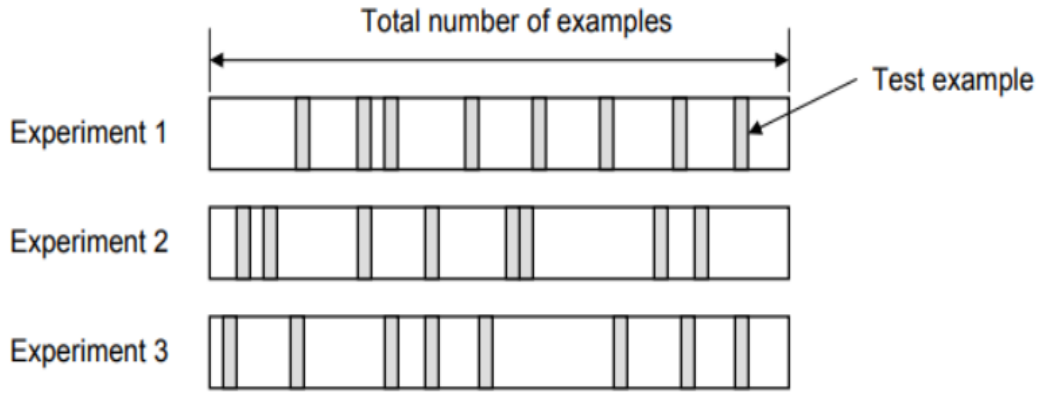

Figure 4.3 Random Subsampling

4.3 Implementation

In this section, our network executed on [32] dataset and its results are presented and compared with several pre-trained networks and the comparison. The machine with high properties used to implement the software Card name: Intel(R) HD Graphics 620, Processor: Intel(R) Core (TM) i7-7500U CPU @ 2.70GHz (4 CPUs), ~2.9GHz Memory: 8192MB RAM with Operating System: Windows 10 Pro 64-bit. The implementation done by using five pre-trained executed and compared with our network:

- AlexNet is the CNN name, presented by Alex Krizhevsky, which trained to classify 1000 classes. This network used in many studies and presented remarkable results. In this study the AlexNet presented 97.26% accuracy which is average of five experiments with 5-fold cross validation see Figure 4.4, Figure 4.5, Figure 4.6, Figure 4.7, Figure 4.8, which represented five confusion matrix for cross 5-fold and finally the average of these results are presented. Furthermore, the average execution time also calculated which presented 2:32 minutes.

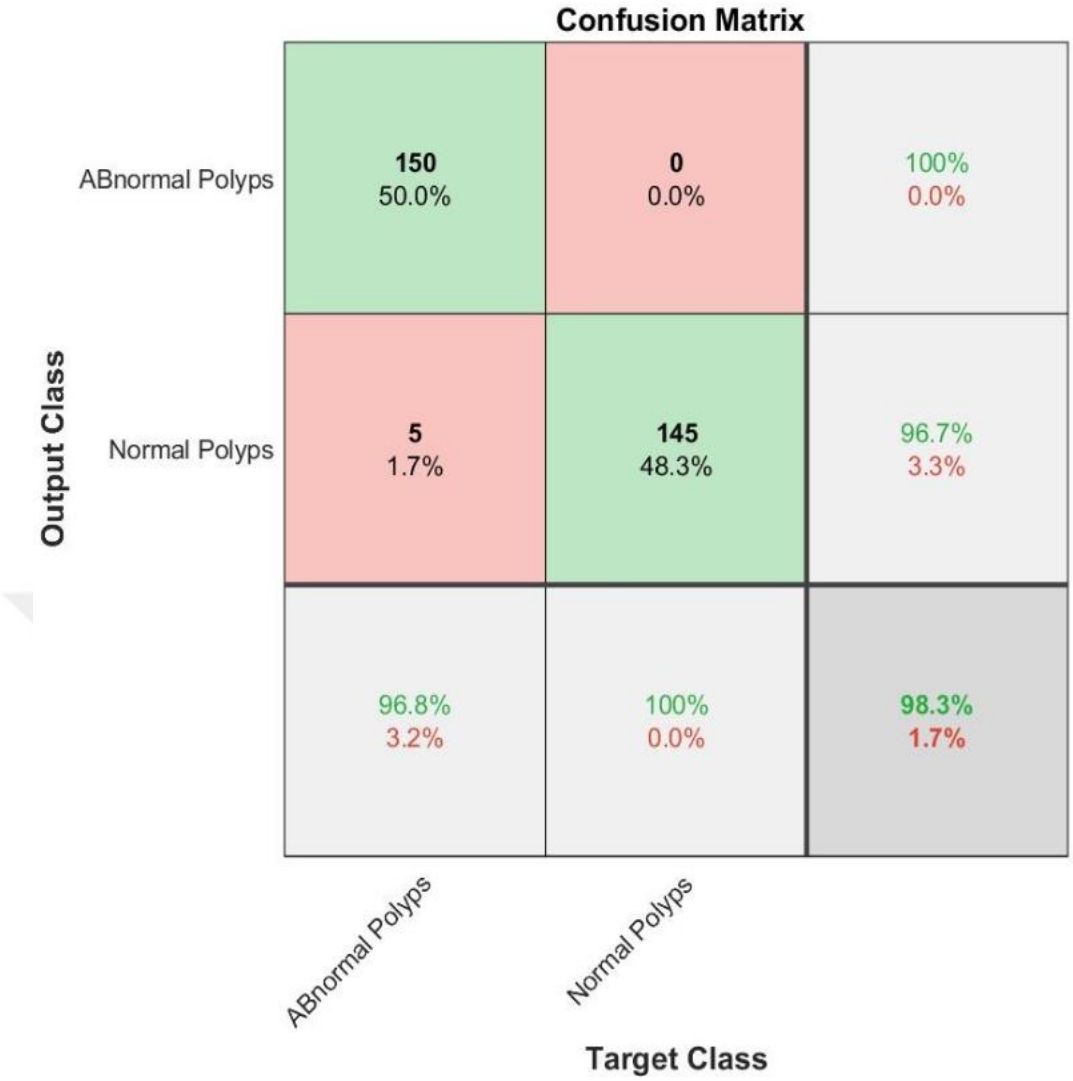

Figure 4.4 AlexNet Results K1

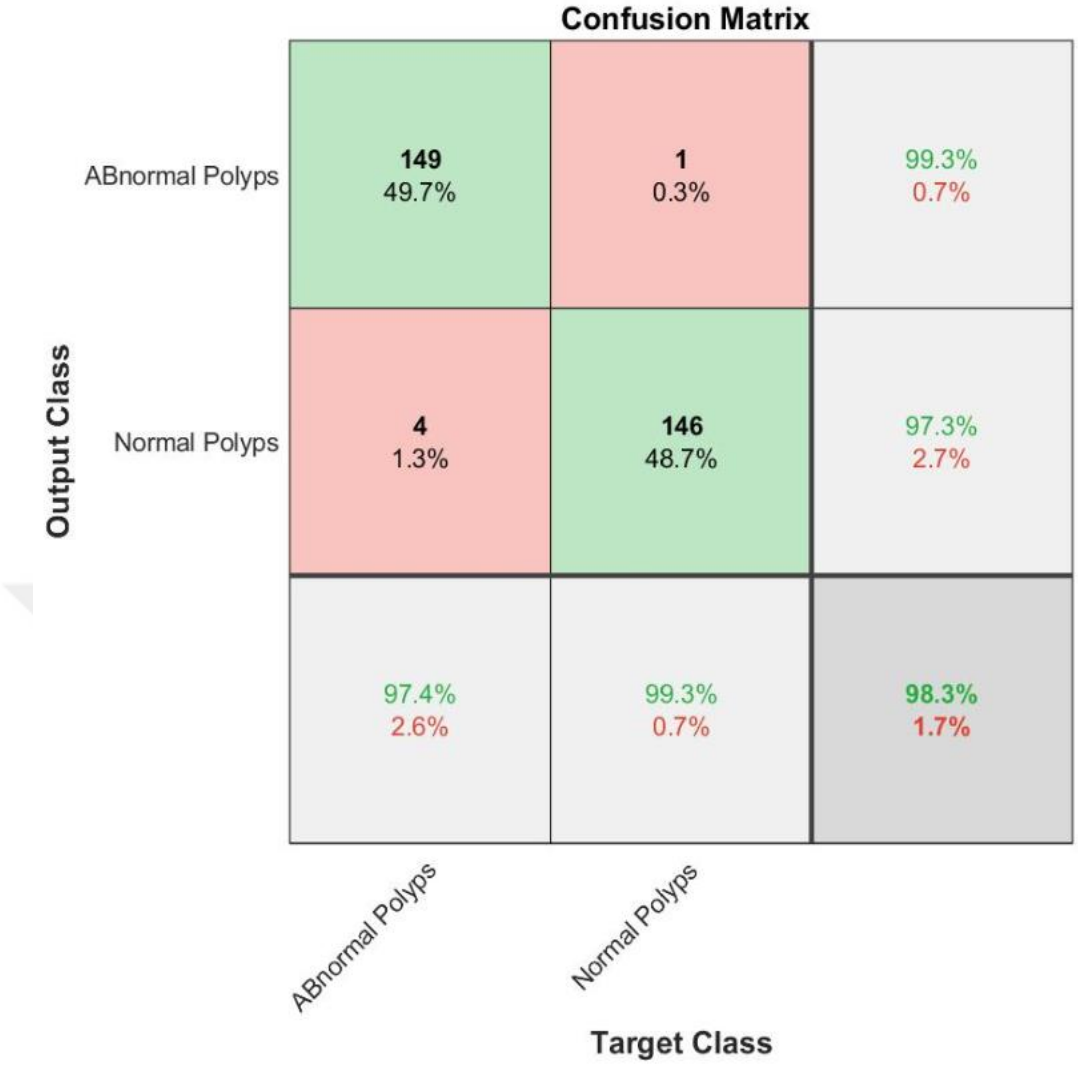

Figure 4.5 AlexNet Results K2

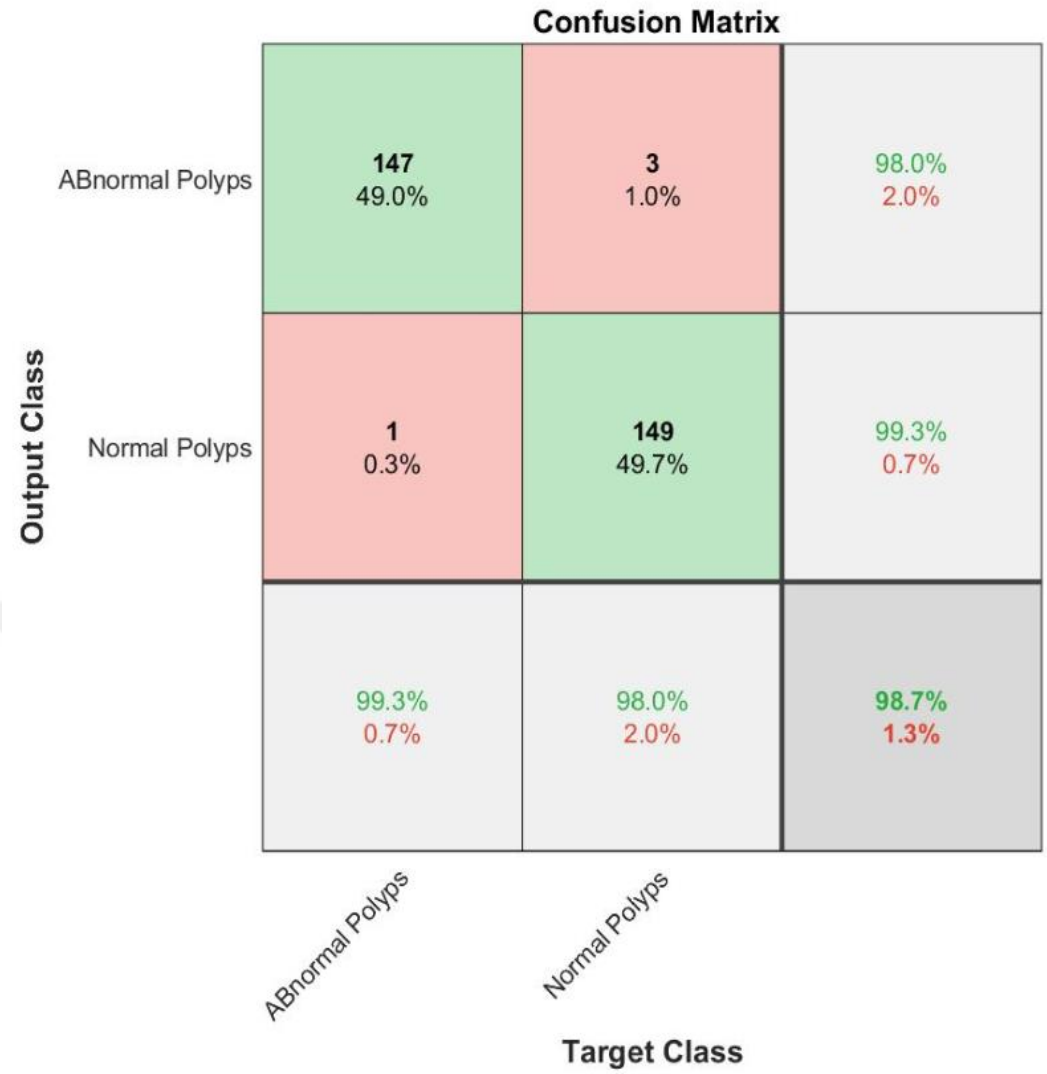

Figure 4.6 AlexNet Results K3

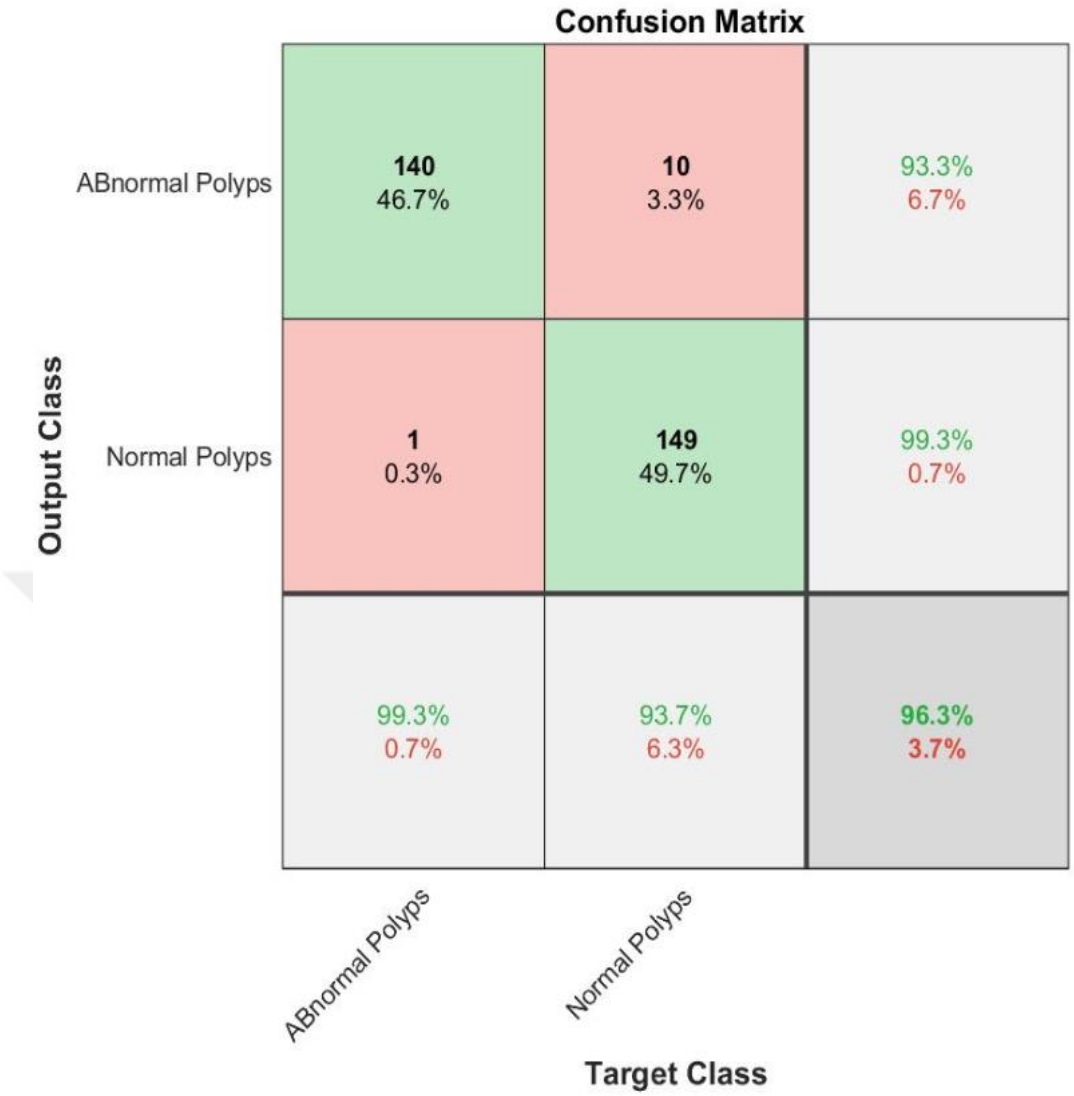

Figure 4.7 AlexNet Results K4

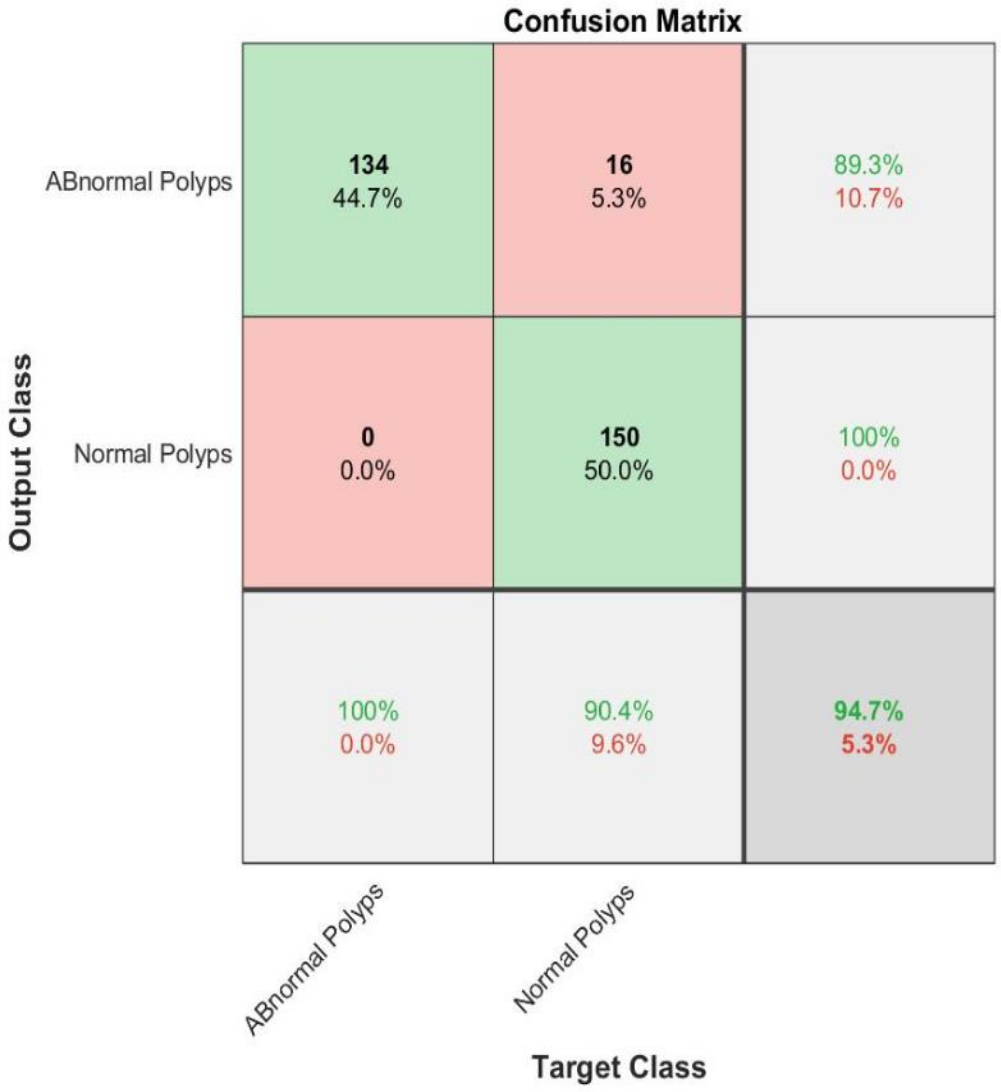

Figure 4.8 AlexNet Results K5

- GoogleNet: GoogleNet is a pre-trained CNN with 22 layers. The network trained to on 1000 class's images such as mouse, pencil and animal. In this study GoogleNet presented 97.68% in 3:44 minutes. The Confusion matrix of GoogleNet presented in Figure 4.9, Figure 4.10, Figure 4.11, Figure 4.12, and Figure 4.13. which represented five confusion matrix for cross 5-fold and finally the average of these results are presented.

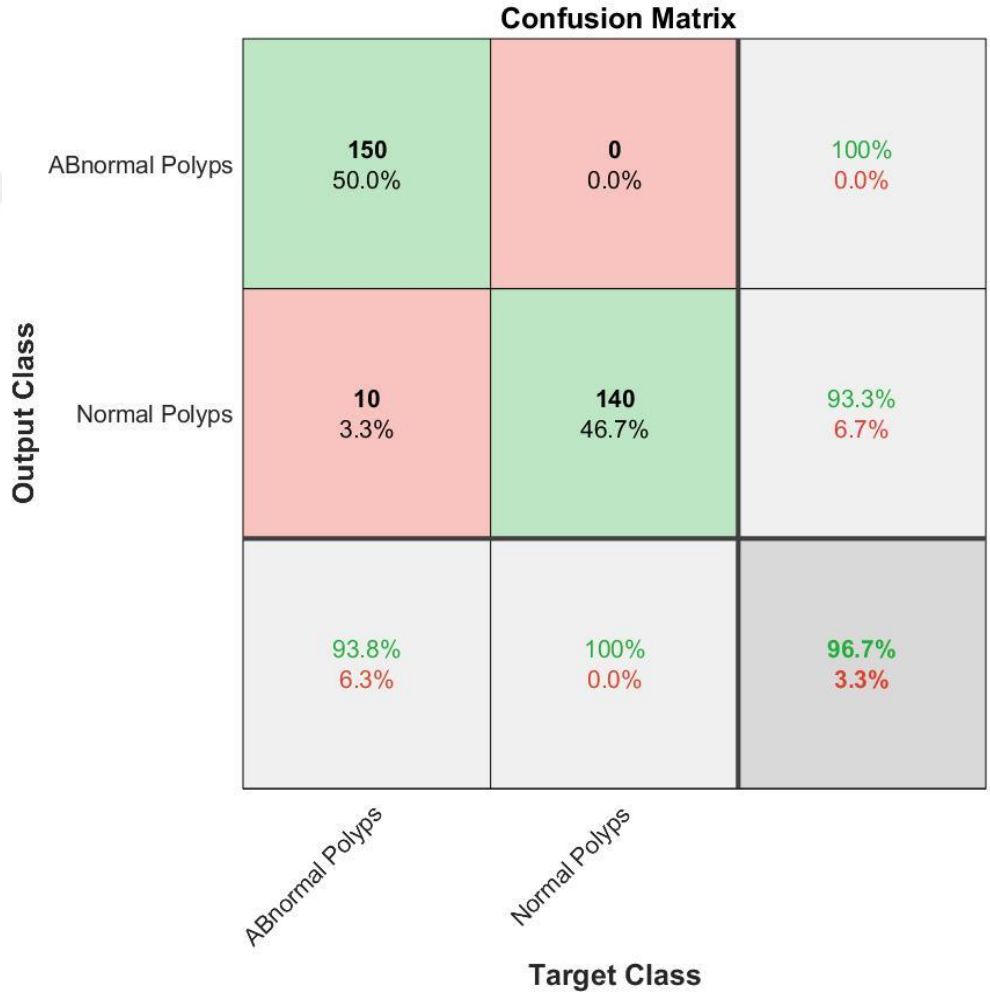

Figure 4.9 GoogleNet Results K1

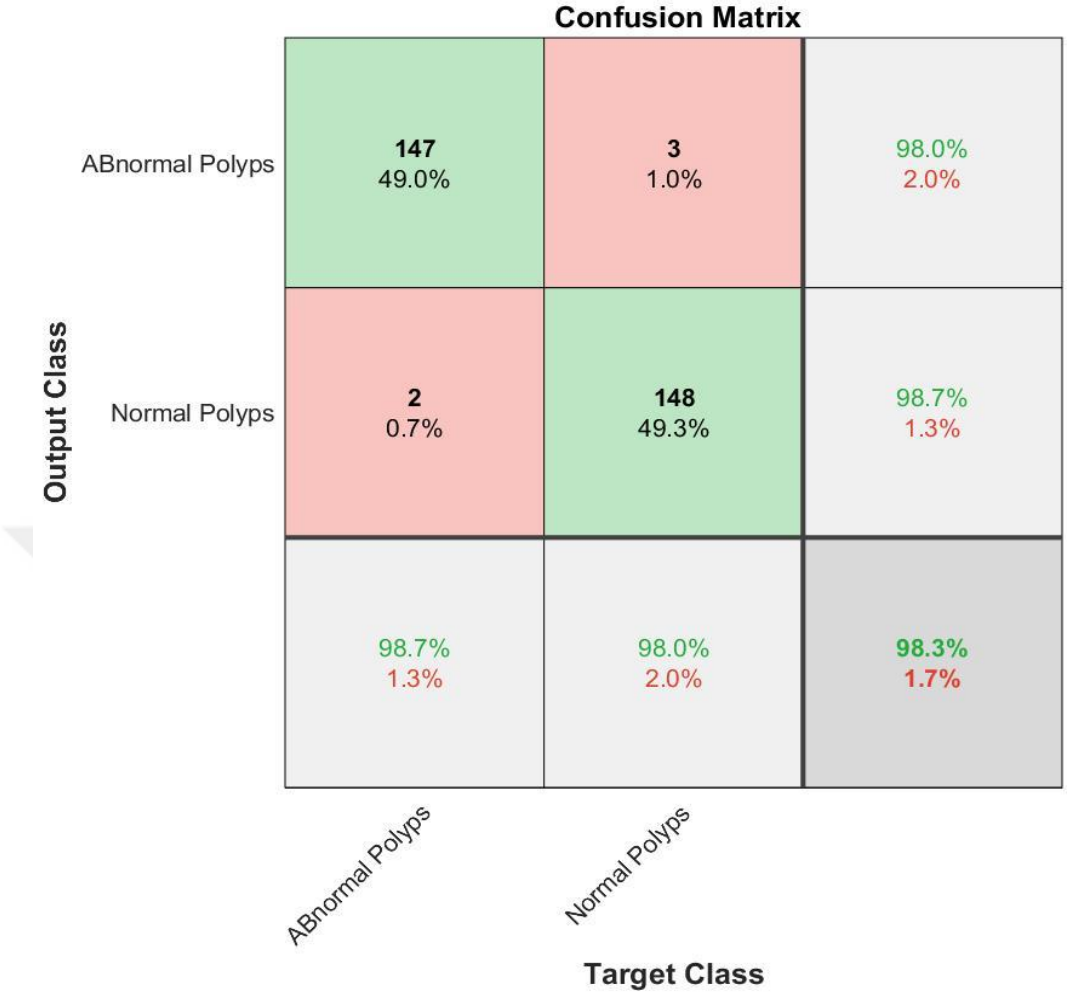

Figure 4.10 GoogleNet Results K2

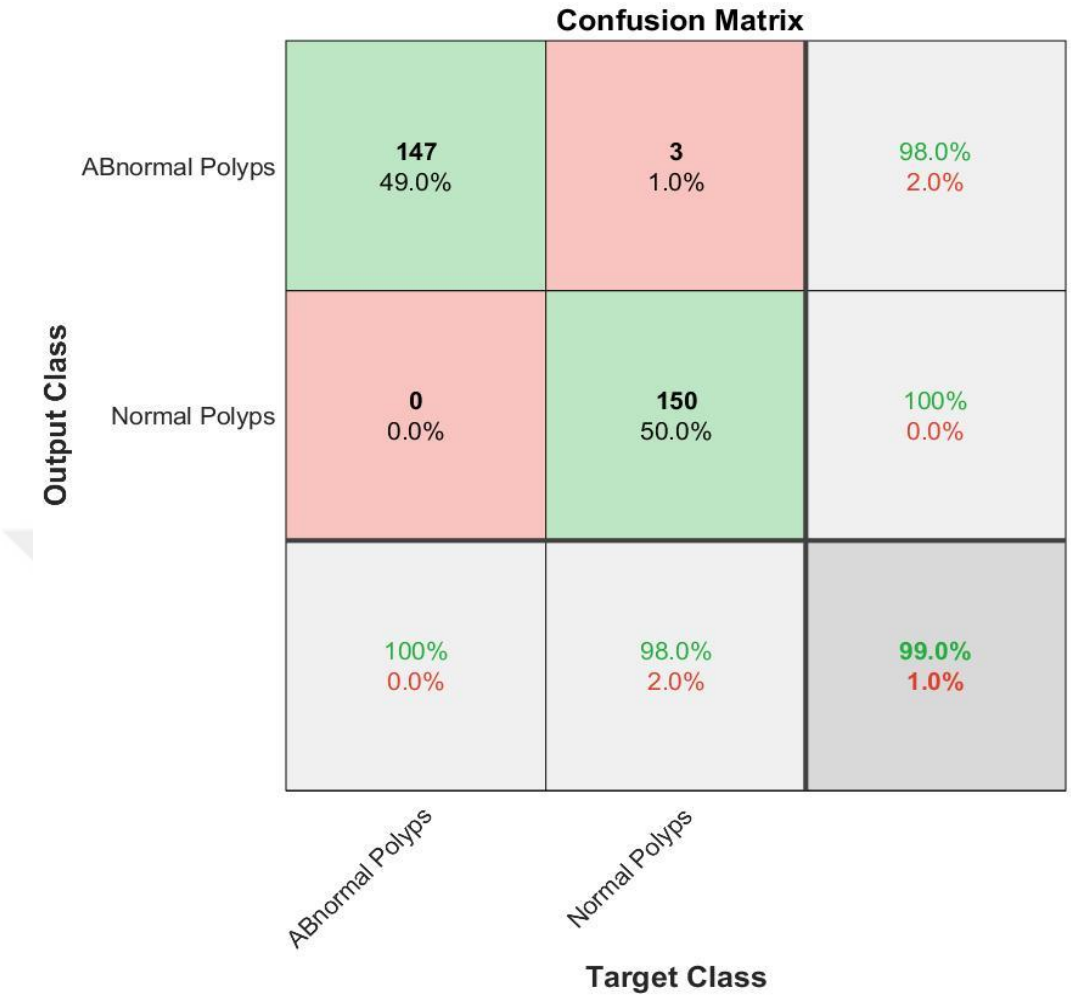

Figure 4.11 GoogleNet Results K3

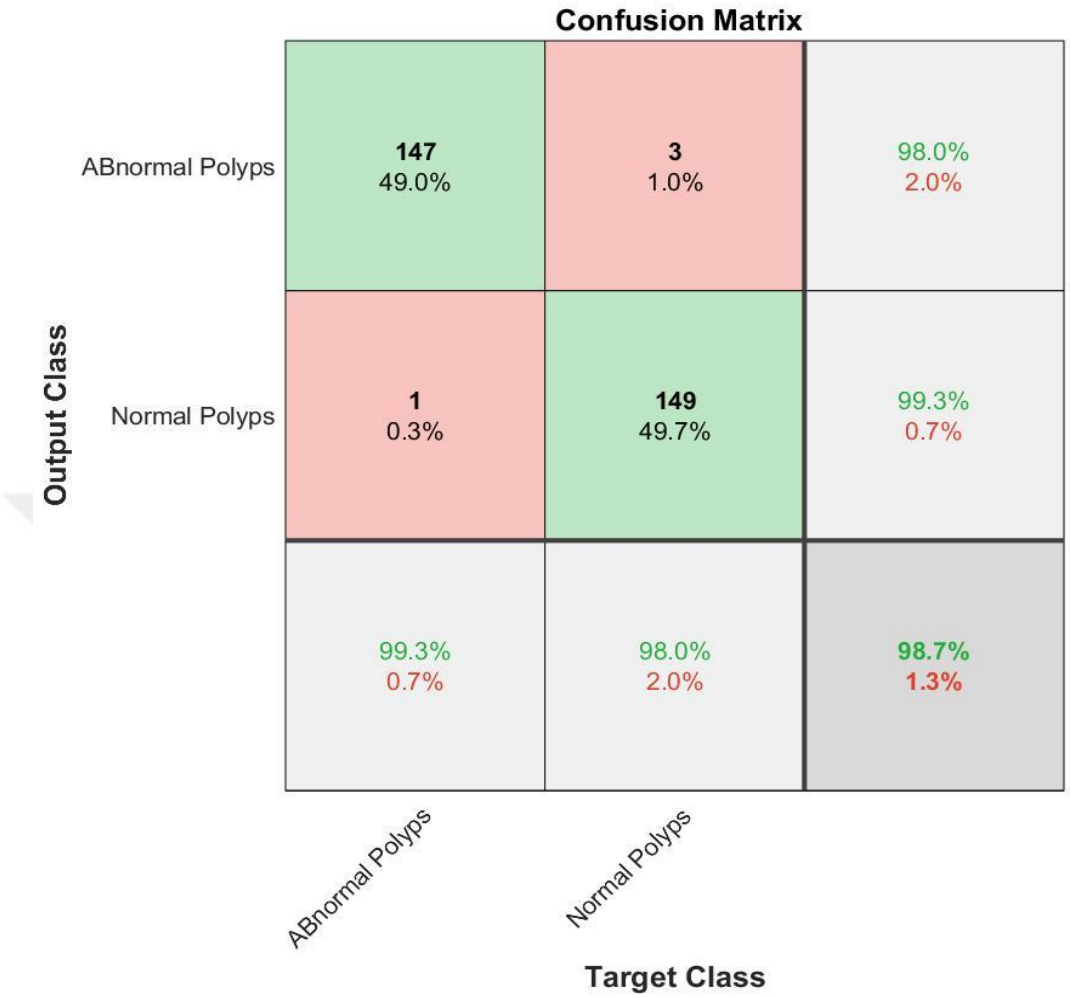

Figure 4.12 GoogleNet Results K4

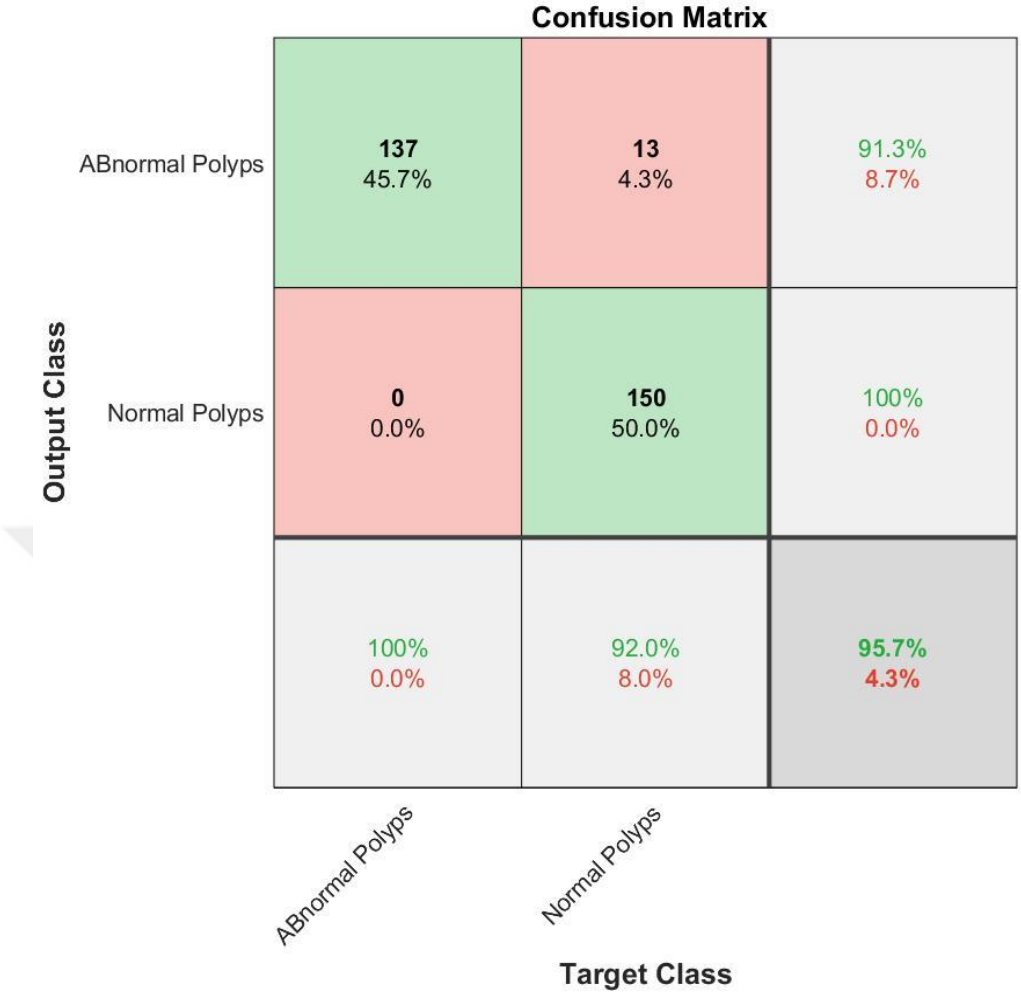

Figure 4.13 GoogleNet Results K5

- ResNet101: ResNet101 is pre-trained CNN trained by using million image and consists from 101 layers. The network trained to classify 1000 objects such as mouse, keyboard and pencil. In this study, pre-trained presented 99.60% in 6:09 minutes which is the average of five experiments accuracies. See Figure 4.14, Figure 4.15, Figure 4.16, Figure 4.17, and Figure 4.18 that presented the ResNet101 Confusion Matrix. which represented five confusion matrix for cross 5-fold and finally the average of these results are presented.

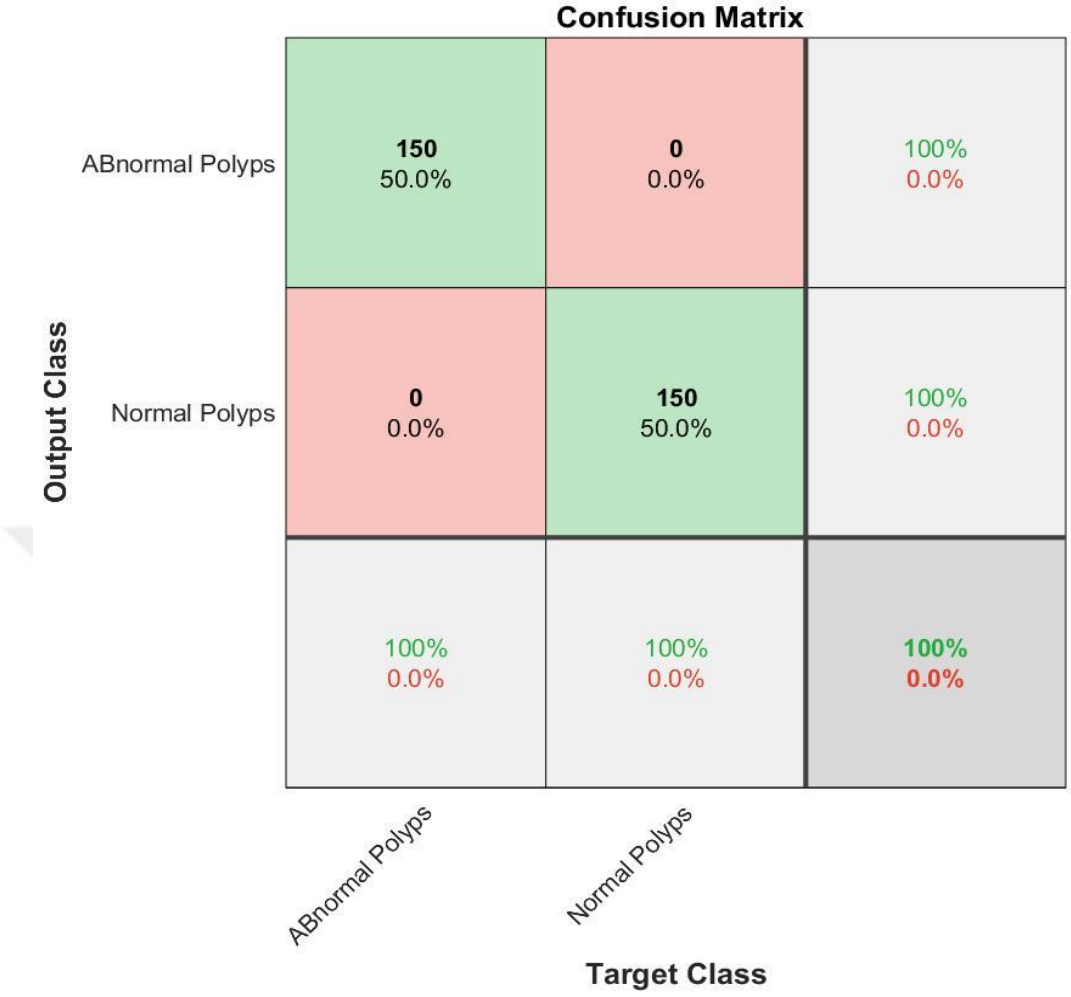

Figure 4.14 ResNet101 Results K1

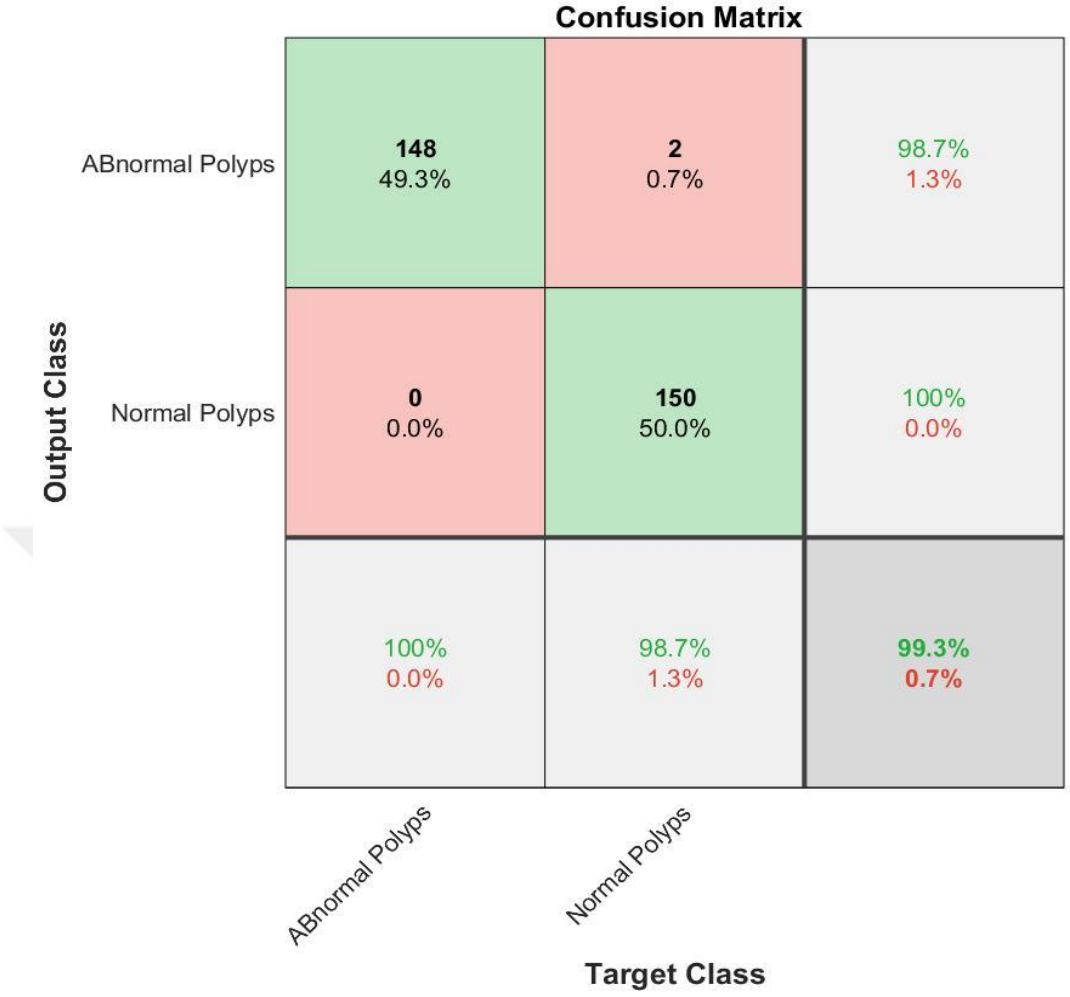

Figure 4.15 ResNet101 Results K2

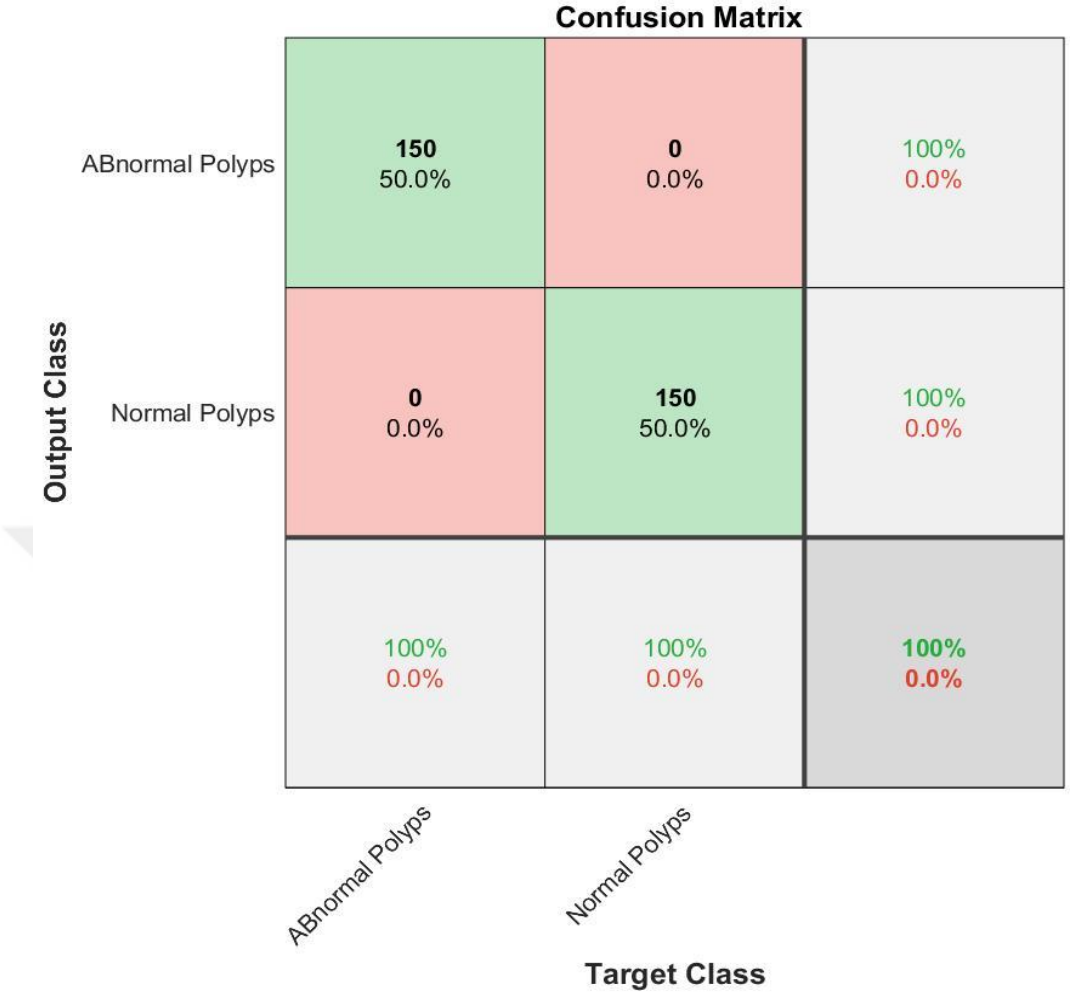

Figure 4.16 ResNet101 Results K3

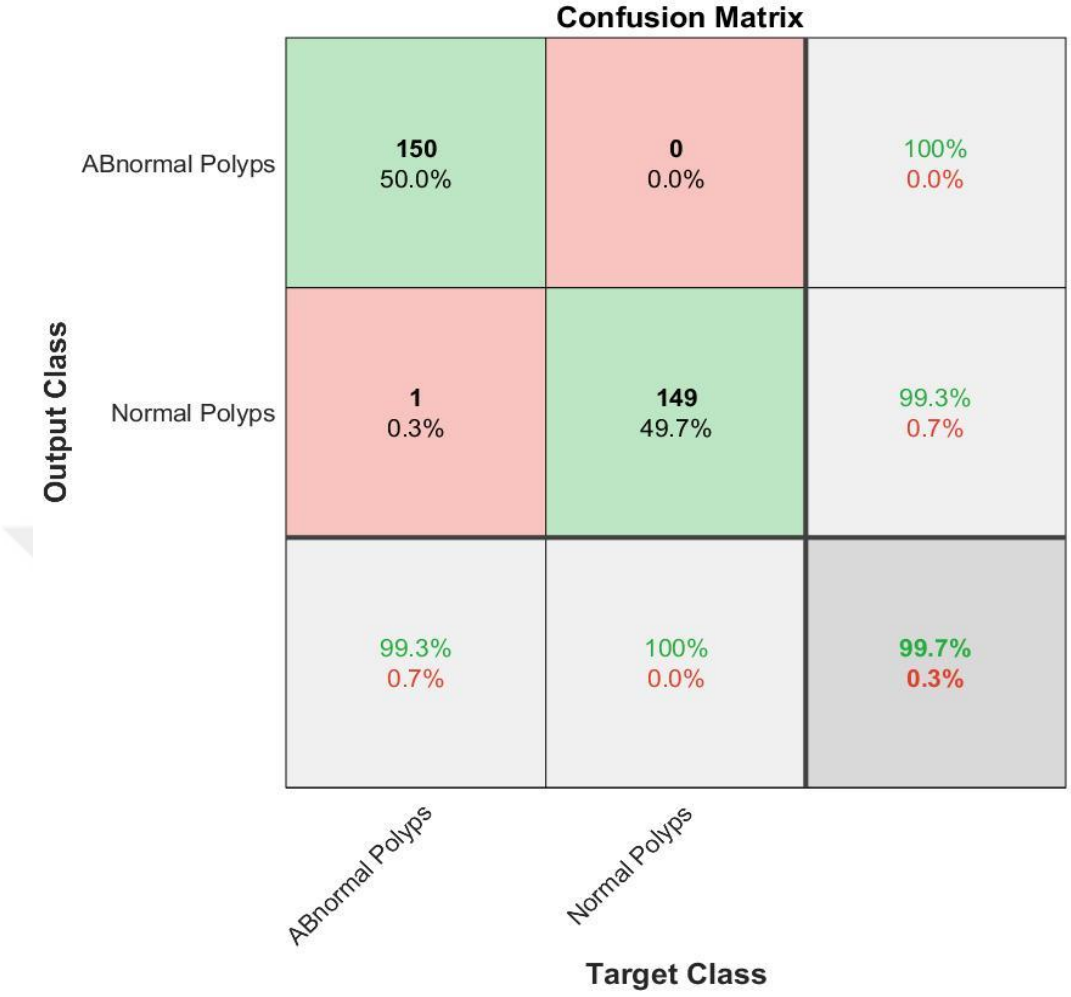

Figure 4.17 ResNet101 Results K4

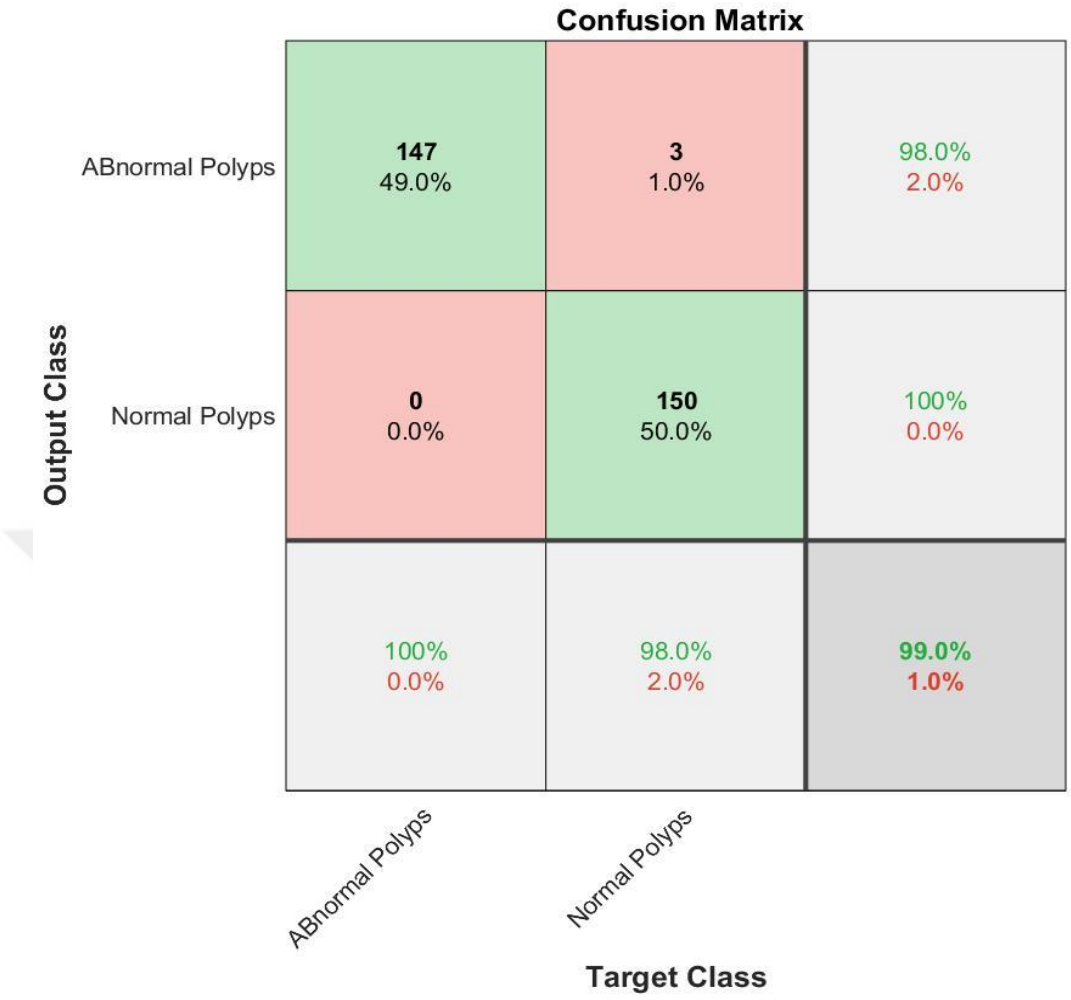

Figure 4.18 ResNet101 Results K5

- DenseNet-201: DenseNet-201 is a CNN that is trained on ImageNet dataset with more than a million images. Furthermore, this network learns to classify 1000 classes and learn important features and consist from 201 layers. This network applied to various studies in different fields. In this study, the DenseNet-201 presented 99.6% with 10:46 minutes for average execution time and which is the average of five experiments accuracies. See Figure 4.19, Figure 4.20, Figure 4.21, Figure 4.22, and Figure 4.23 which represented five confusion matrix for cross 5-fold and finally the average of these results are presented.

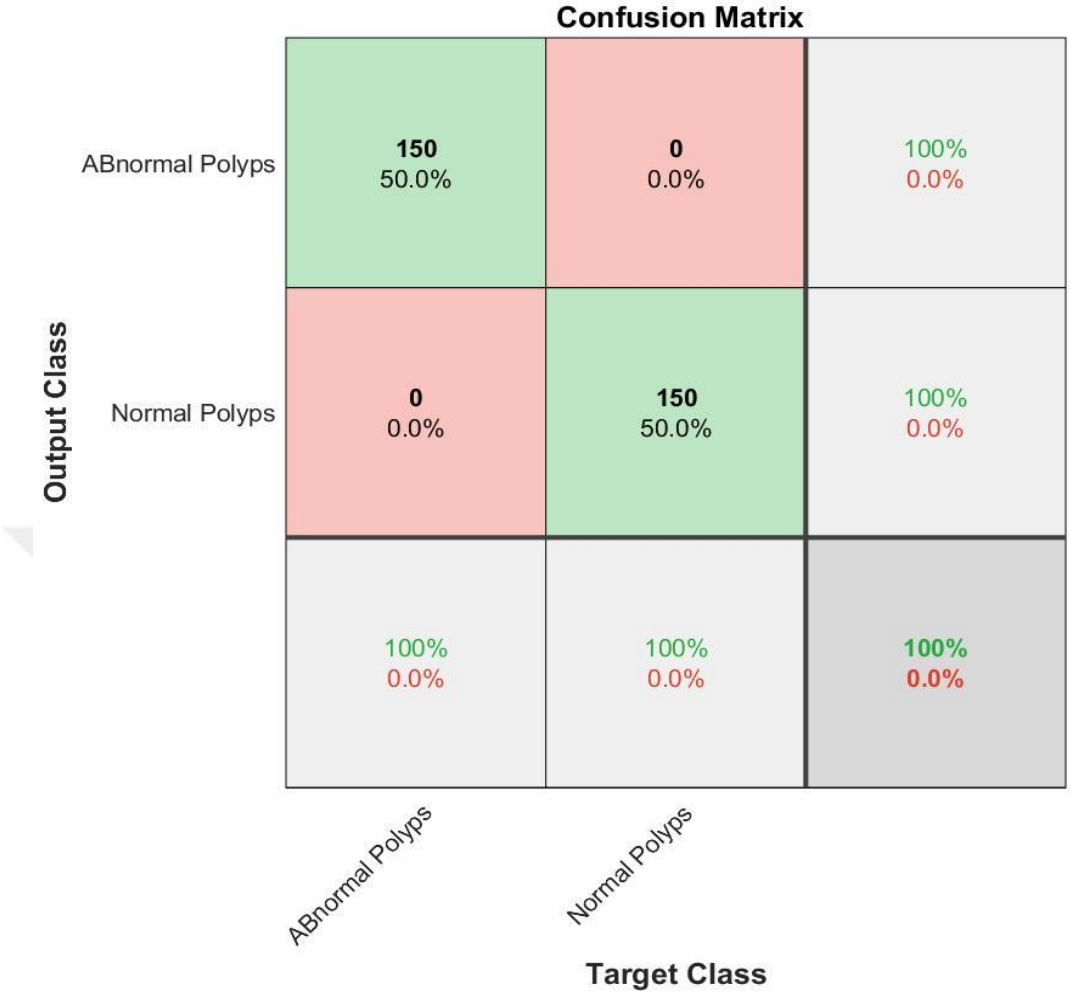

Figure 4.19 DenseNet 201 Results K1

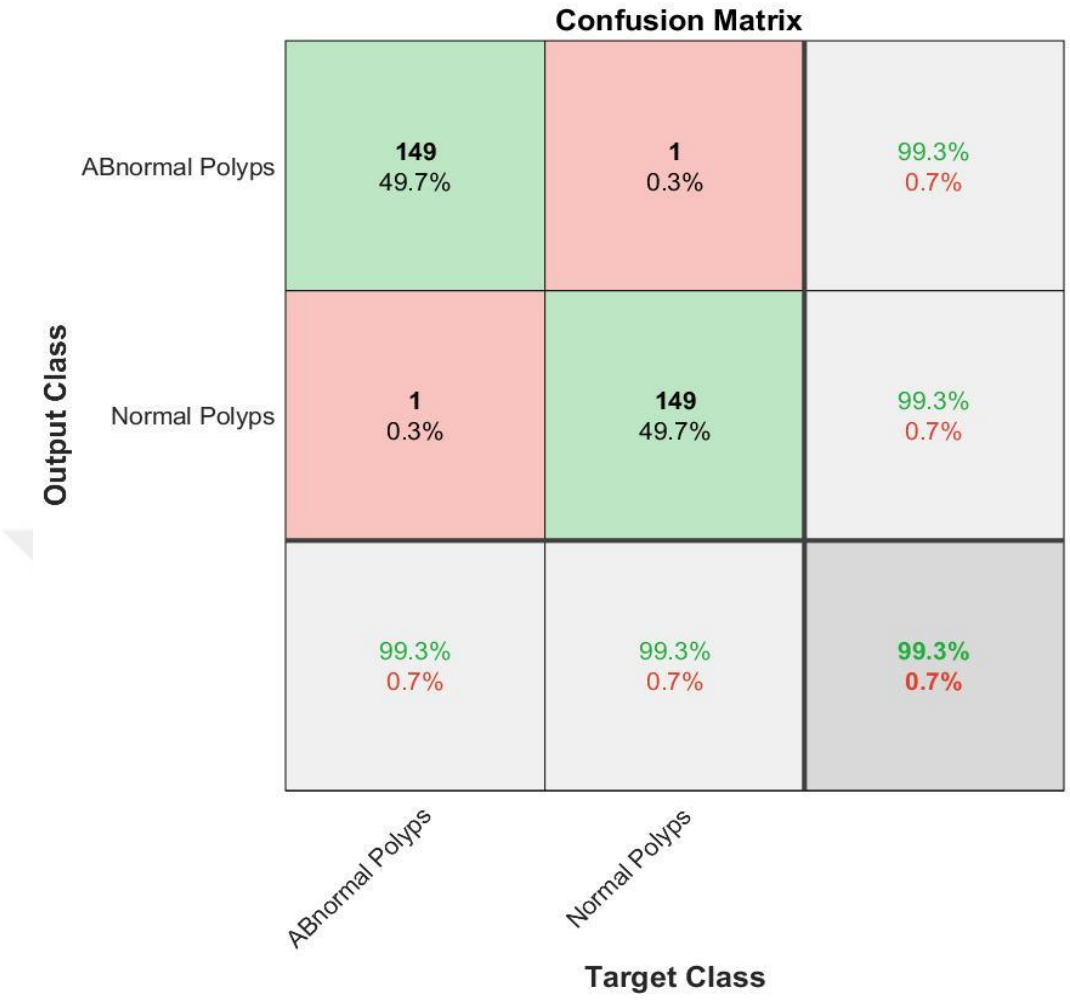

Figure 4.20 DenseNet 201 Results K2

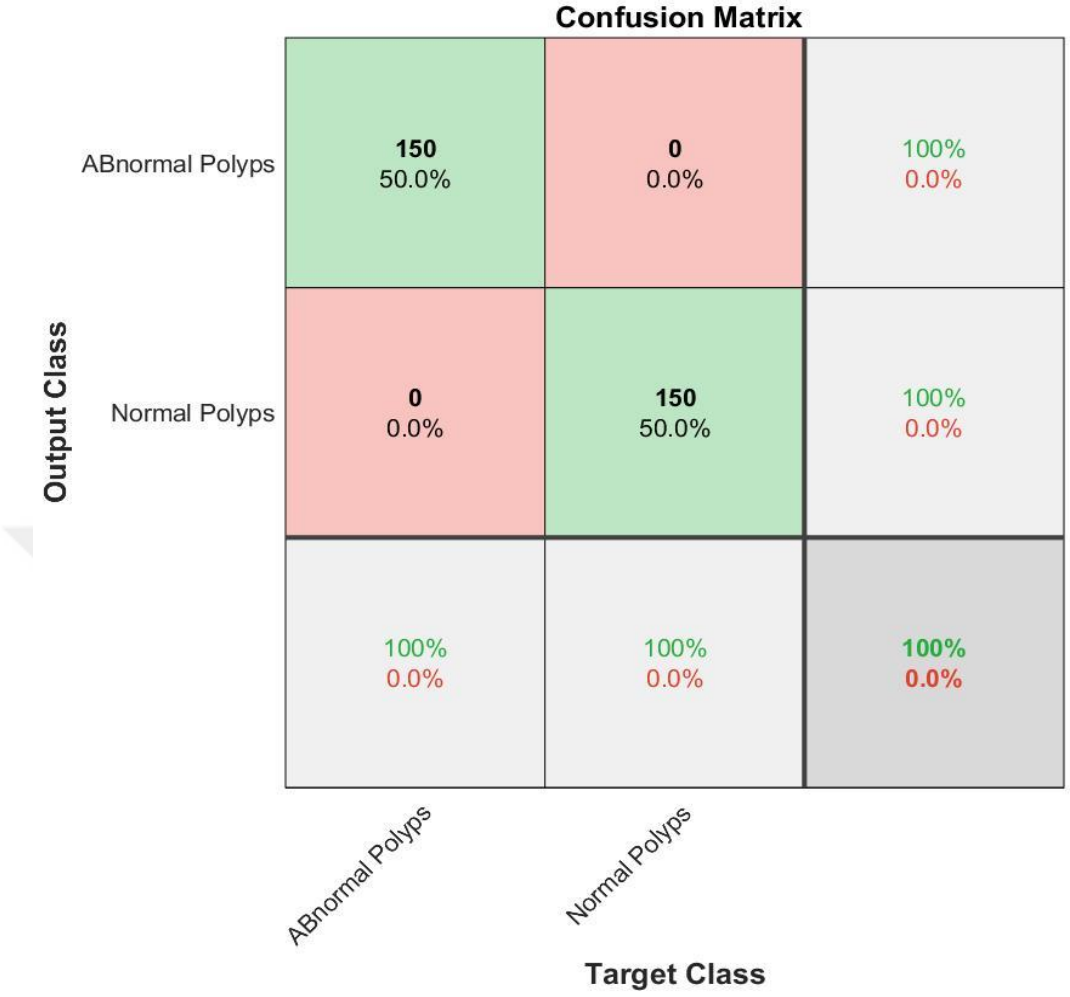

Figure 4.21 DenseNet 201 Results K3

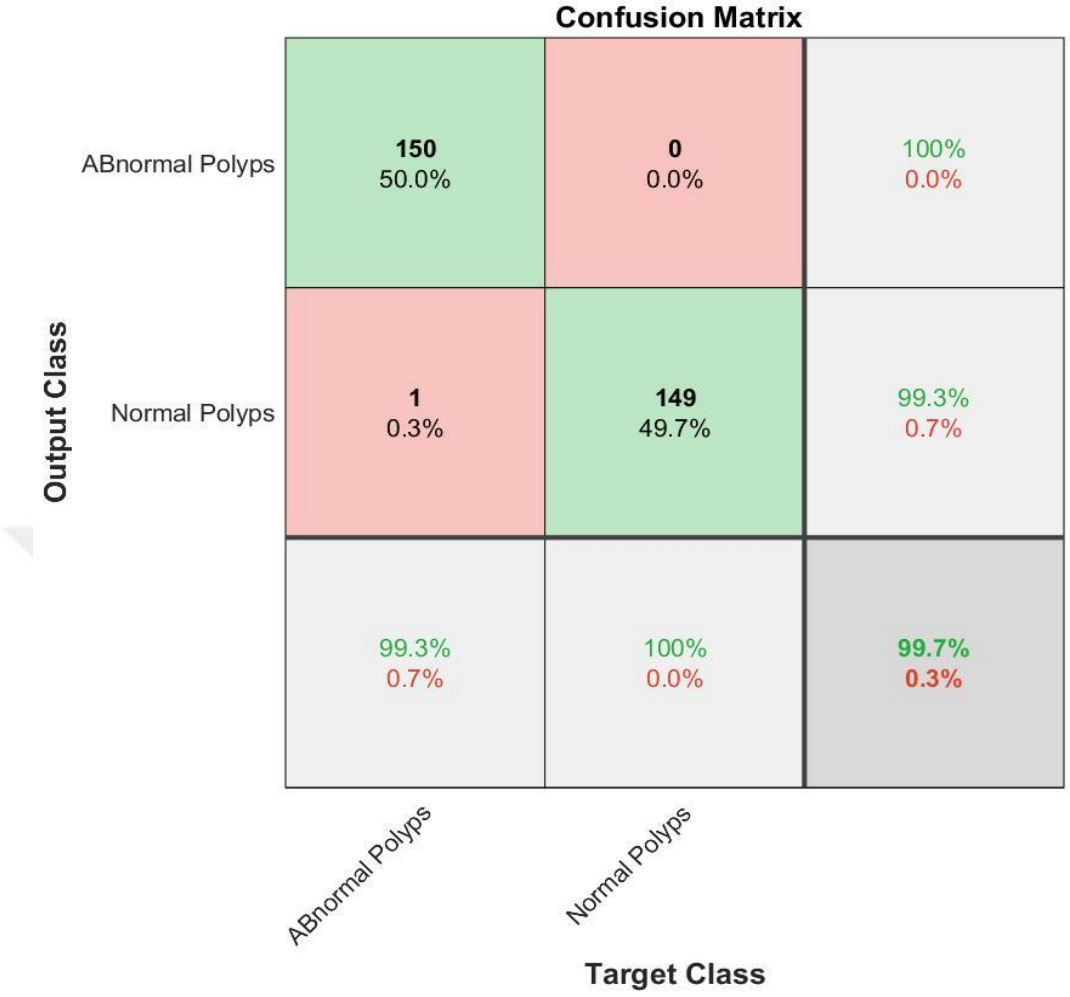

Figure 4.22 DenseNet 201 Results K4

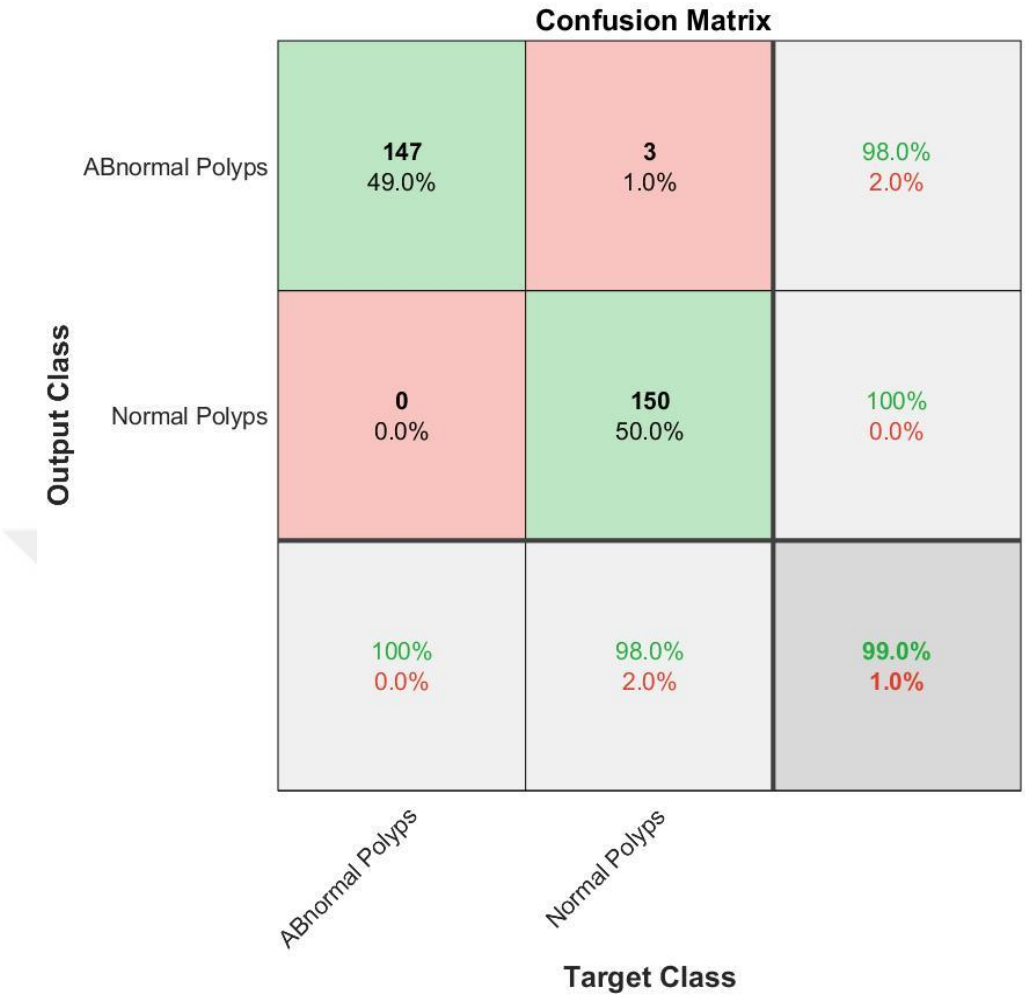

Figure 4.23 DenseNet 201 Results K5

- VGG16: VGG16 is a CNN that is trained on ImageNet dataset with more than a million images. The VGG16 trained to classify 1000 object and contain from 41 layers. In this study, the VGG16 network presented 98.34% accuracy with 9:20 minutes which is the average of five experiments. See Figure 4.24, Figure 4.25, Figure 4.26, Figure 4.27, Figure 4.28 which represented five confusion matrix for cross 5-fold and finally the average of these results are presented.

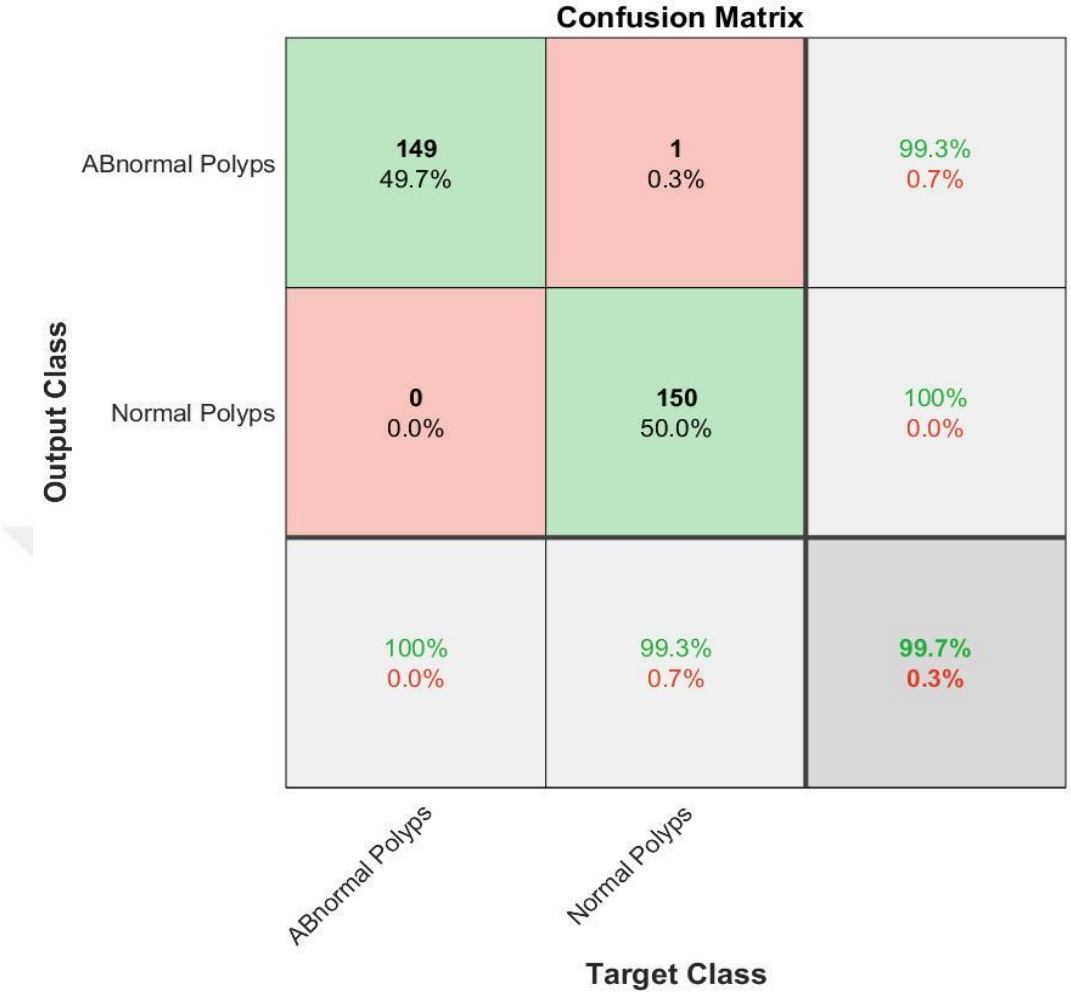

Figure 4.24 VGG16 Results K1

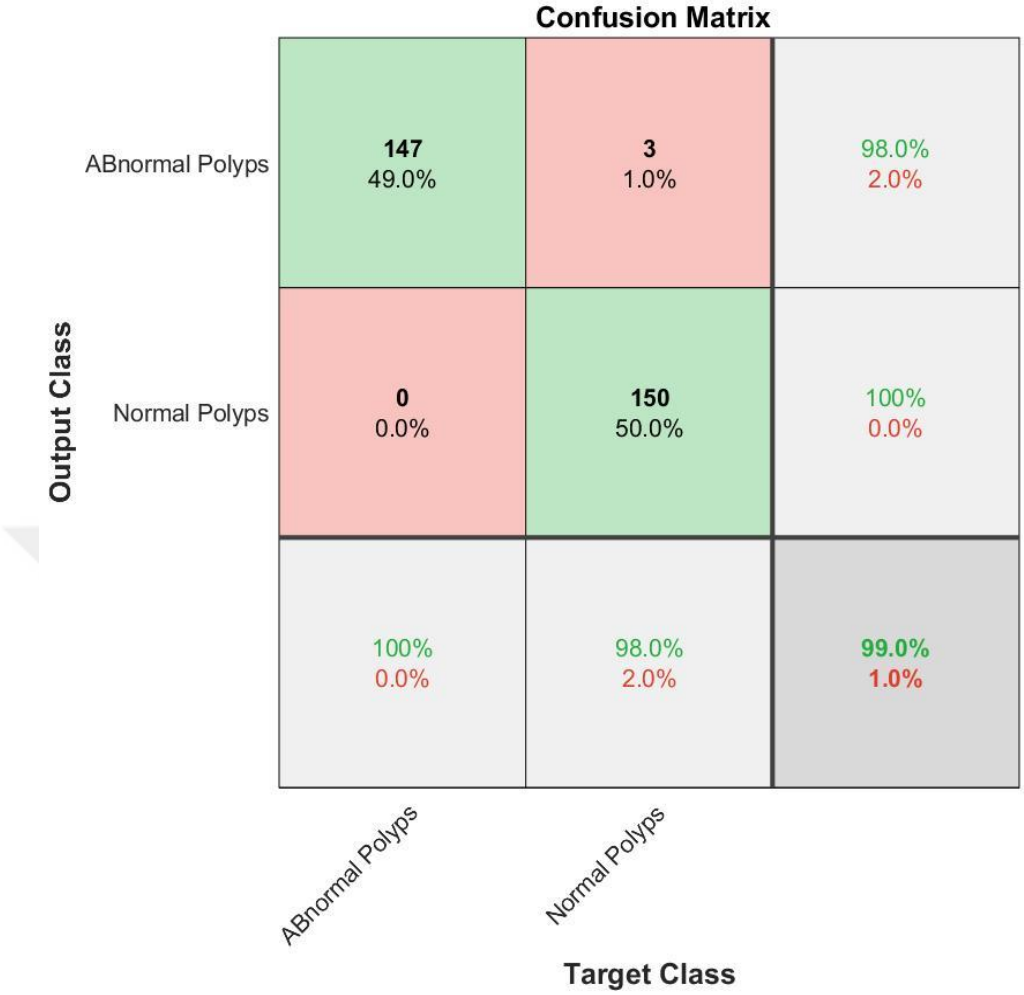

Figure 4.25 VGG16 Results K2

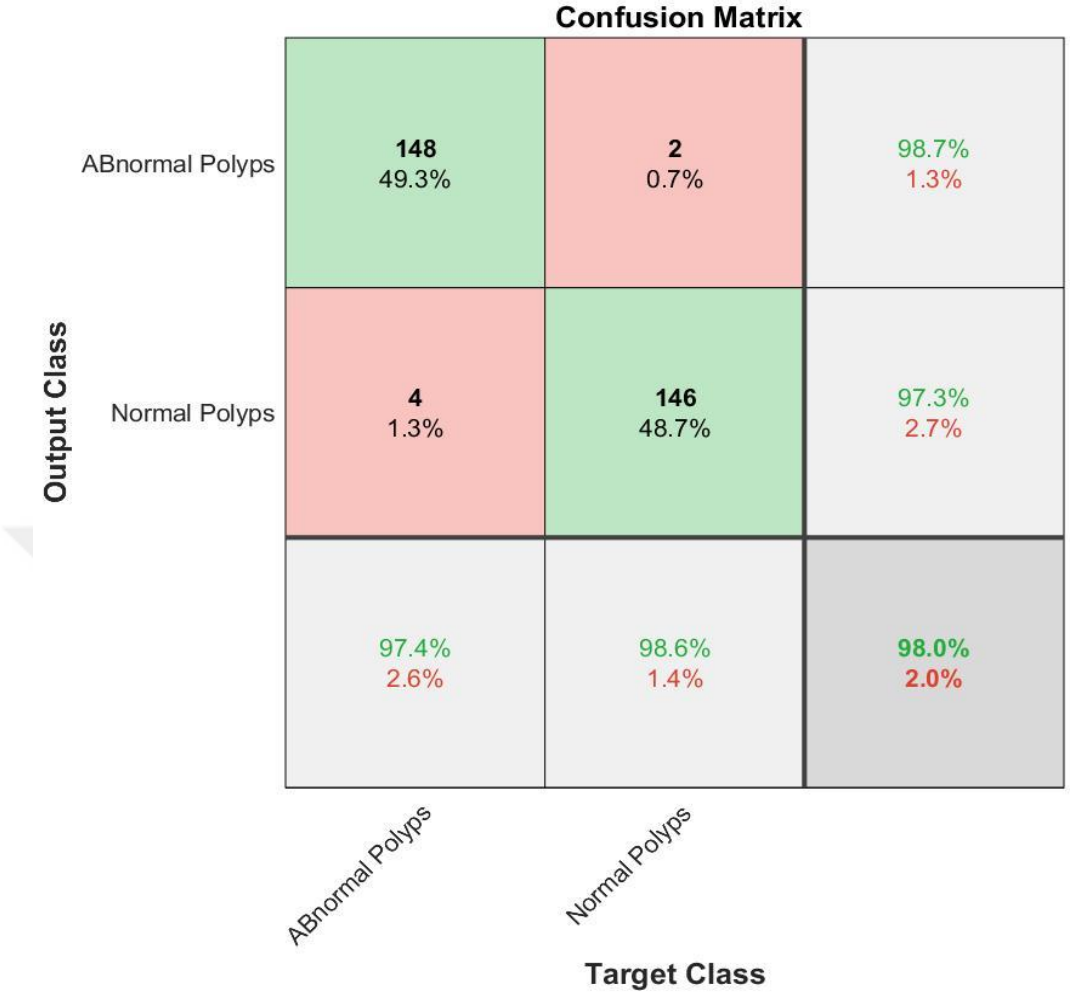

Figure 4.26 VGG16 Results K3

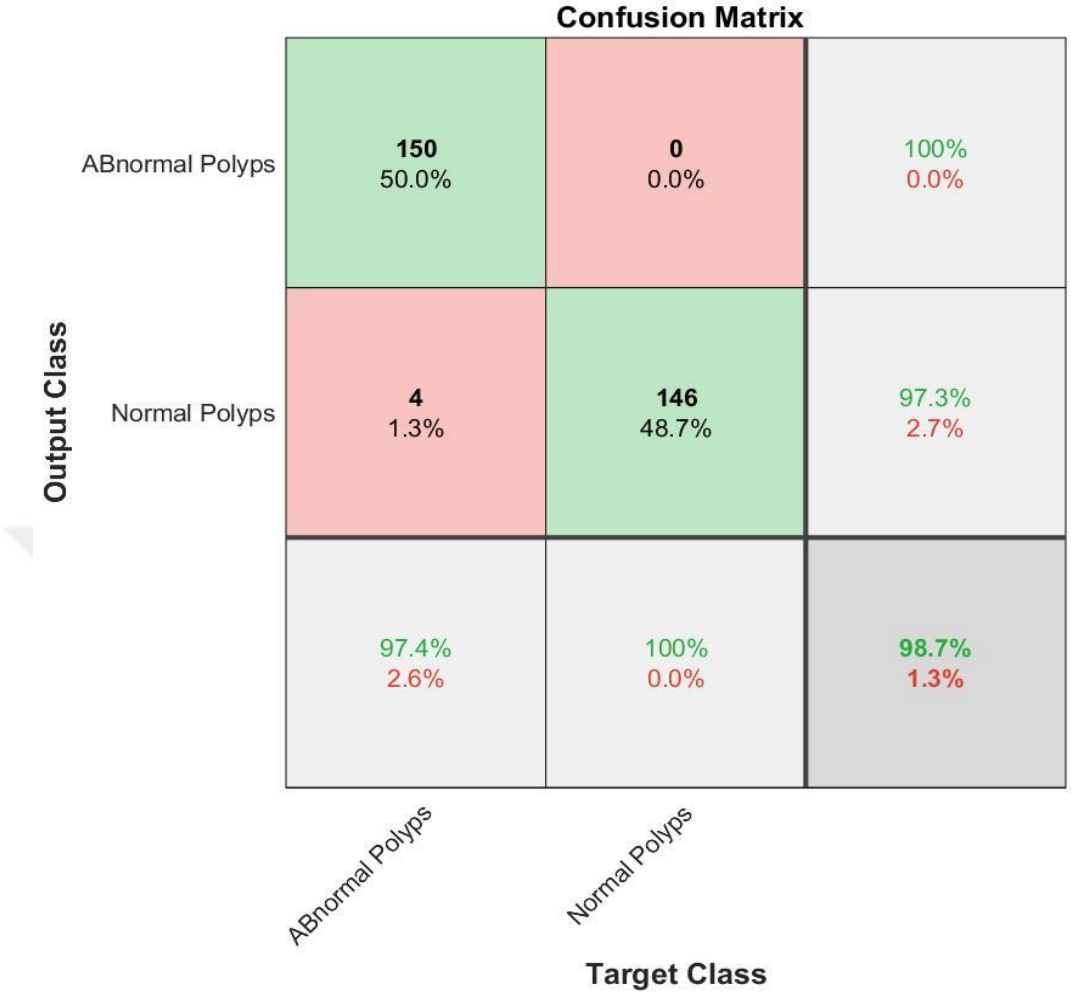

Figure 4.27 VGG16 Results K4

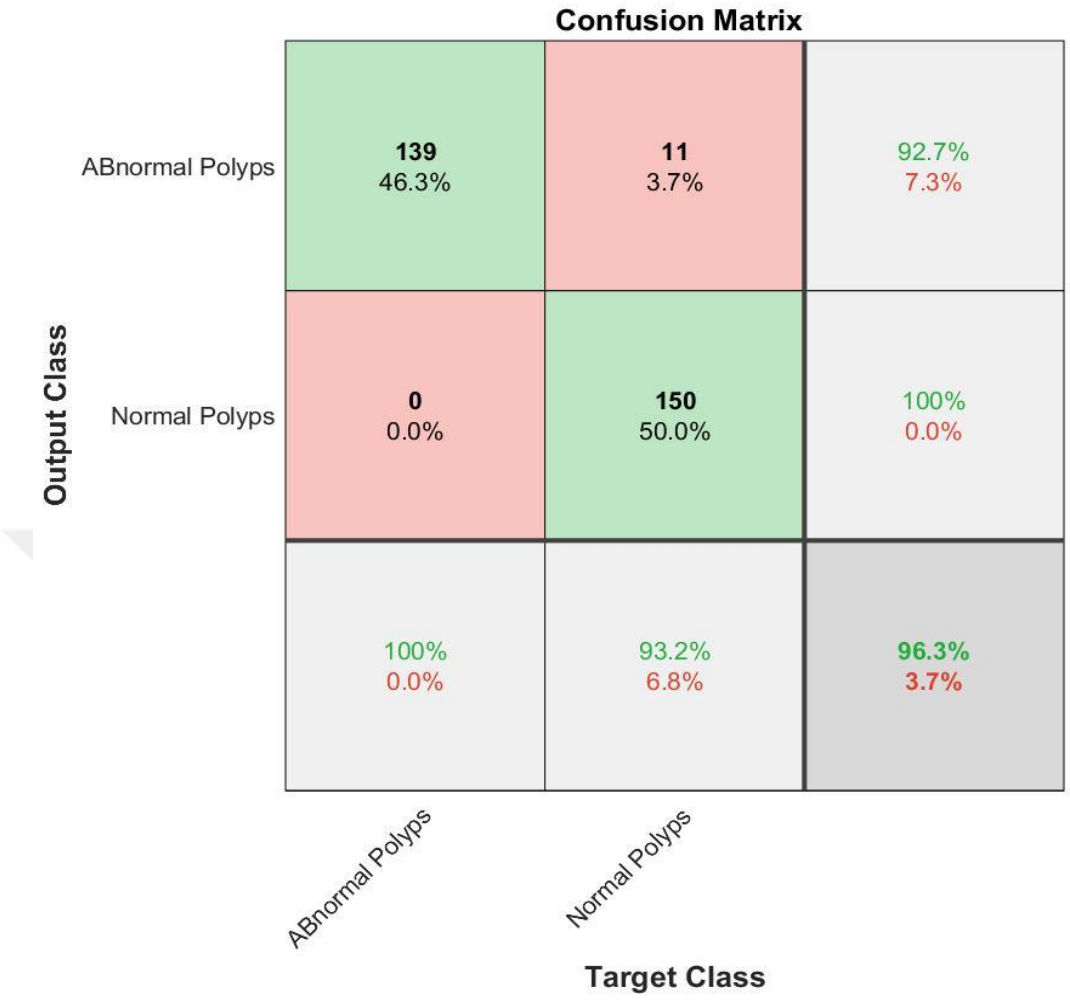

Figure 4.28 VGG16 Results K5

- On other hand, our network trained only with gastric cancer detection to classify the data into two classes and consist from 8 layers. Our network presented 99.88% in 3:05 minutes which is the average of five experiments. Our network presented remarkable results when compared with other pre-trained networks in both accuracy and execution time. See Figure 4.29, Figure 4.30, Figure 4.31, Figure 4.32, and Figure 4.33 that presented our method Confusion Matrix. which represented five confusion matrix for cross 5-fold and finally the average of these results are presented.

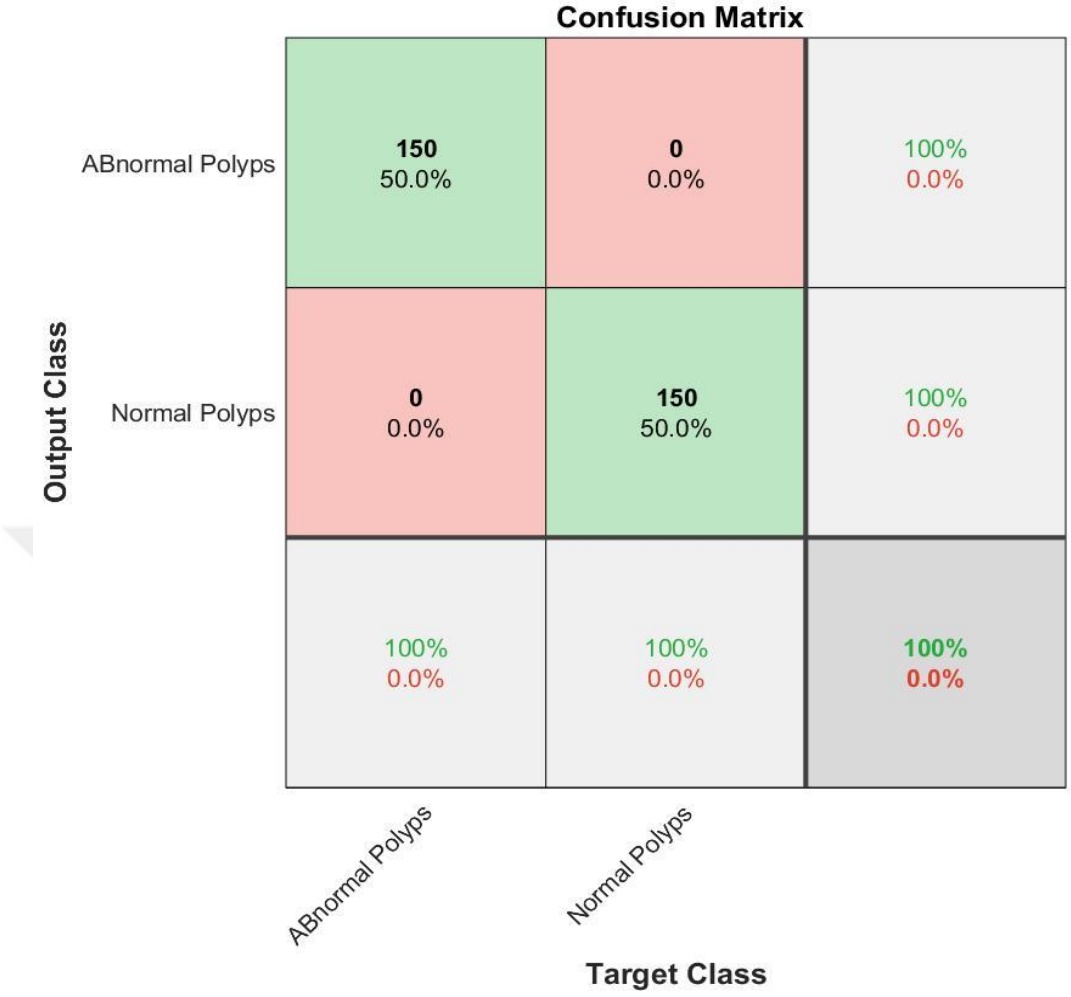

Figure 4.29 Our Method Results K1

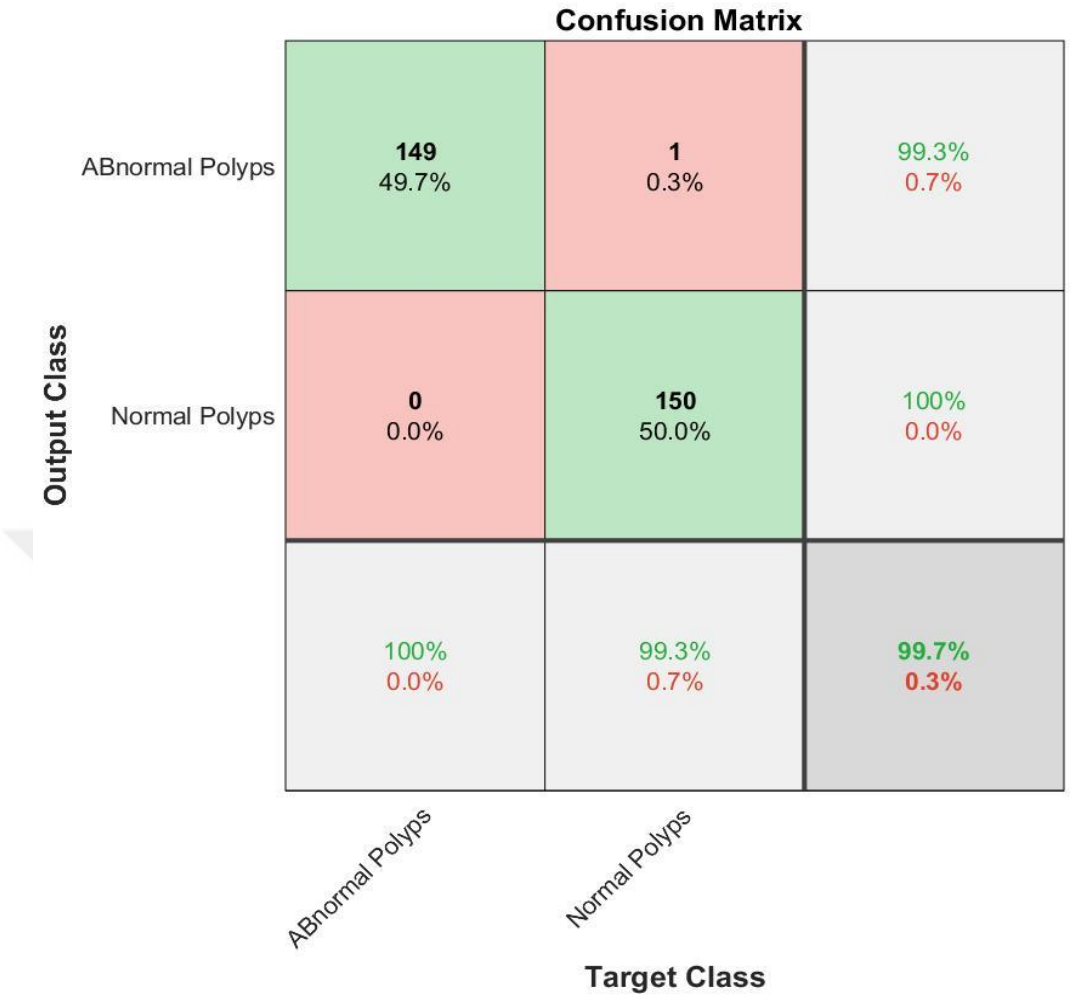

Figure 4.30: Our Method Results K2

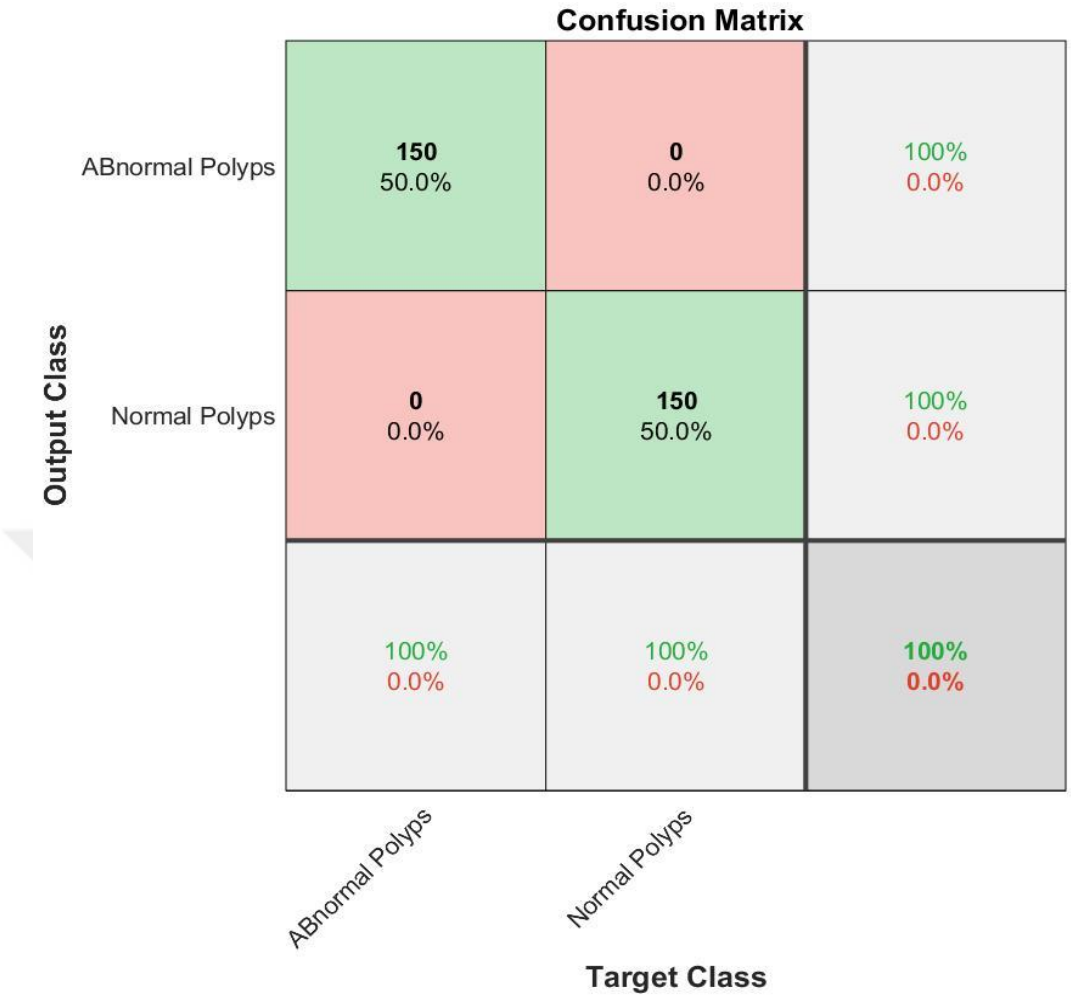

Figure 4.31 Our Method Results K3

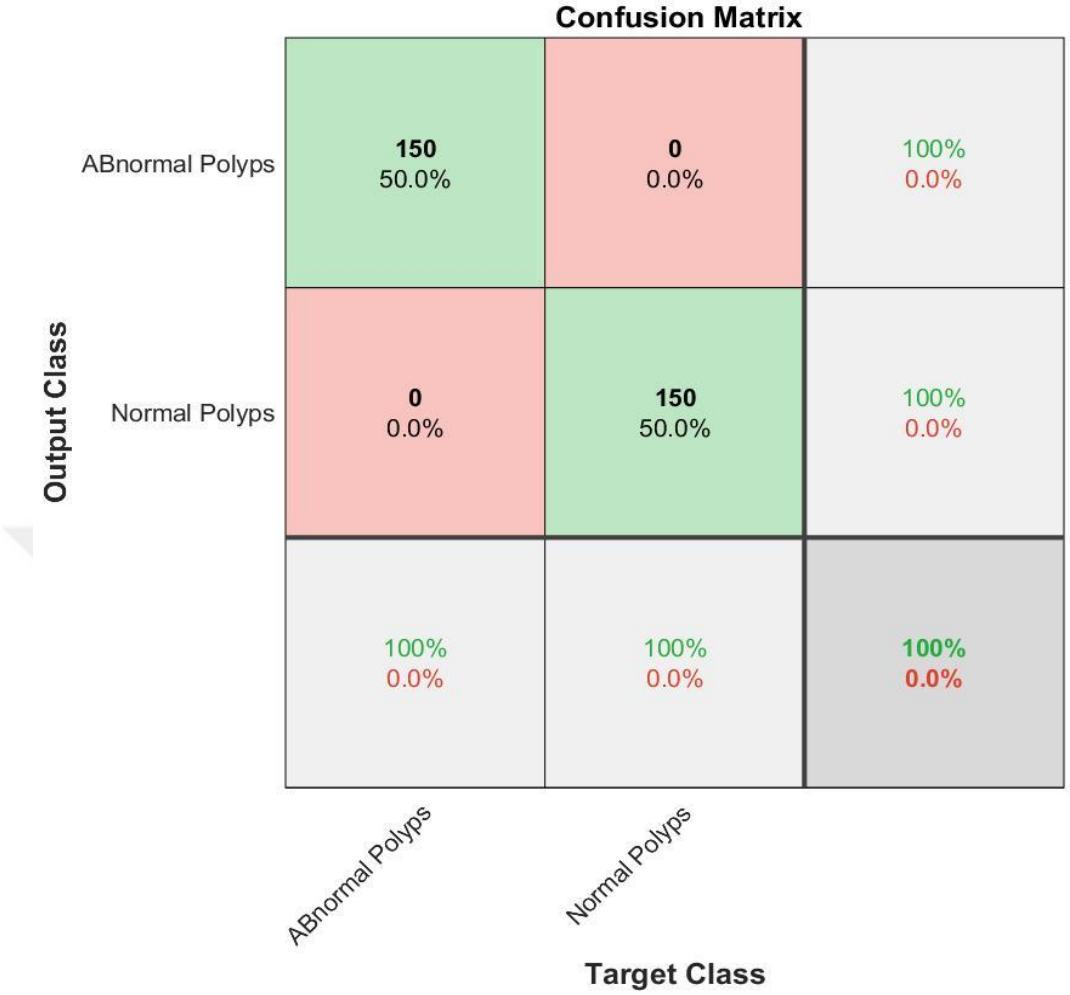

Figure 4.32 Our Method Results K4

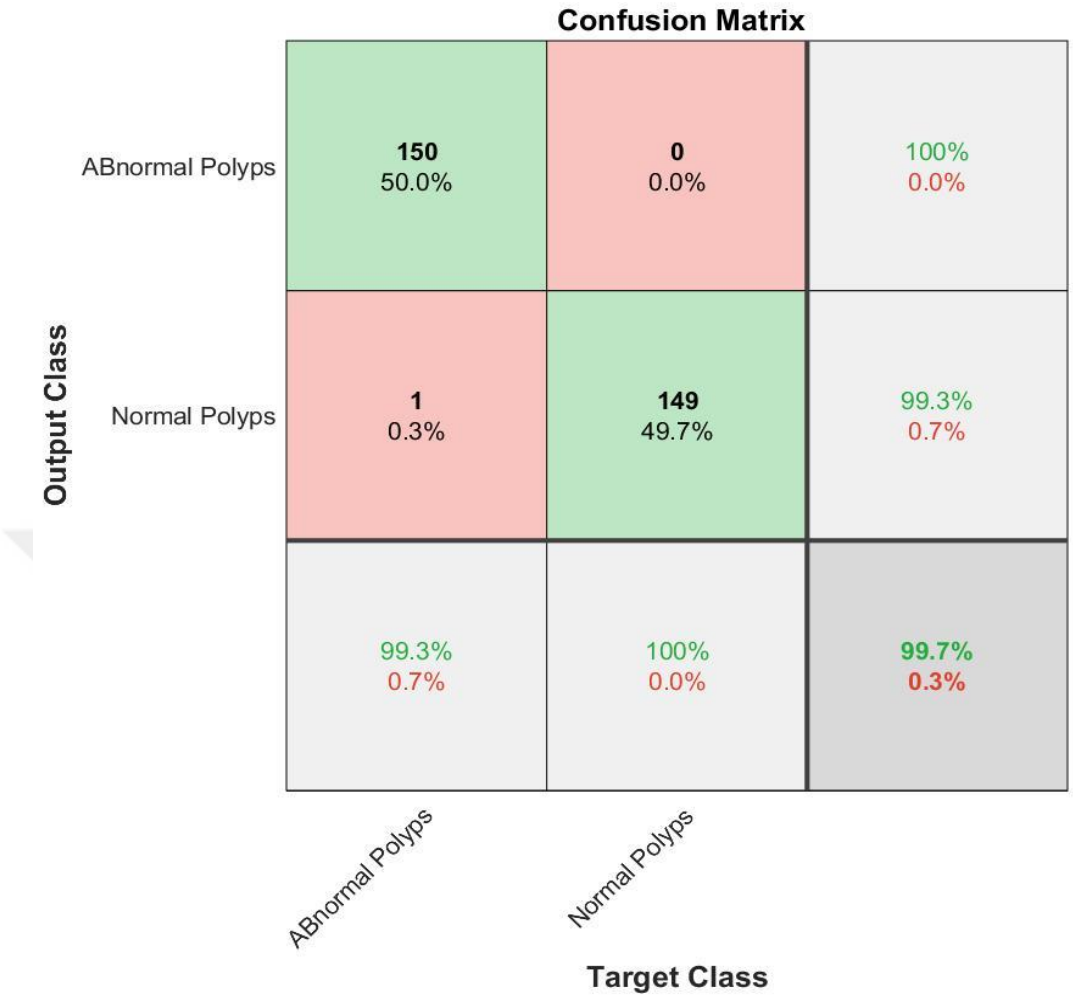

Figure 4.33 Our Method Results K5

In the Table 4.1, the results, execution times, structures of network and accuracies are presented. AlexNet presented the fastest execution time with lowest accuracy and the proposed method presented the highest accuracy but with time execution more than AlexNet. The GoogleNet presented fastest execution time compared with our method but also our method presented high accuracy compared with GoogleNet. On the other hand, our method presented best results compared with five pretrained convolution neural network that also applied to our problem.

Furthermore, the presented results compared with each other's and represented graphically. See the Figure 4.34.

On the other hand, the execution times also compared in represented in the Figure 4.35.

Table 4.2 Comparison Table

Networks	Execution time (Minute)	Accuracy (%)	Image input size (pixels)	Layers for Network
AlexNet	**2:32**	97.26	227 x 227	25
GoogleNet	3:44	97.68	224 x 224	144
ResNet101	6:09	99.60	224 x 224	347
VGG16	9:20	98.33	224 x 224	41
DenseNet201	10:46	99.60	224 x 224	77
OurNet	3:05	**99.88**	224 x 224	8

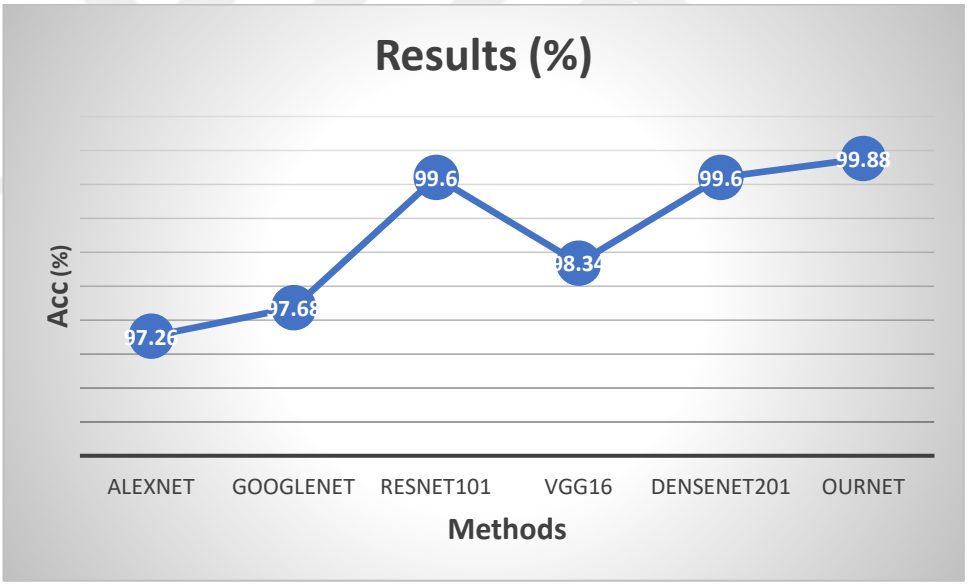

Figure 4.34 Results Comparision

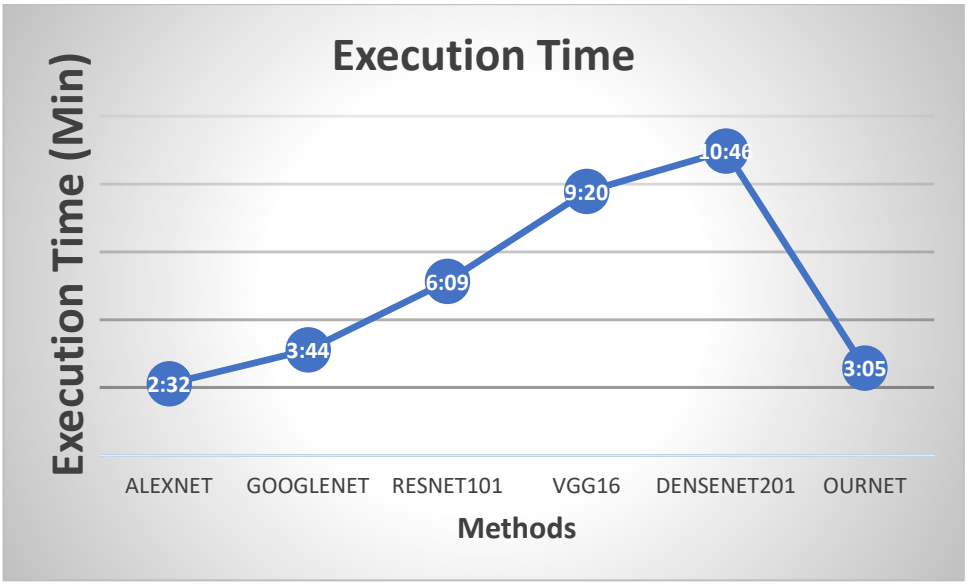

Figure 4.35 Execution Times Comparision

On the other hand several layers of CNN are tested to prove which number of layers presented the best results with SoftMax classifier.

In the first stage, choosing one layer of CNN with SoftMax classifier directly with five K-Fold cross validation are executed and the average of it is presented in bold in the table 4.3 and these results for each K are represneted in figure 4.36, 4.37, 4.38, 4.39 and figure 4.40.

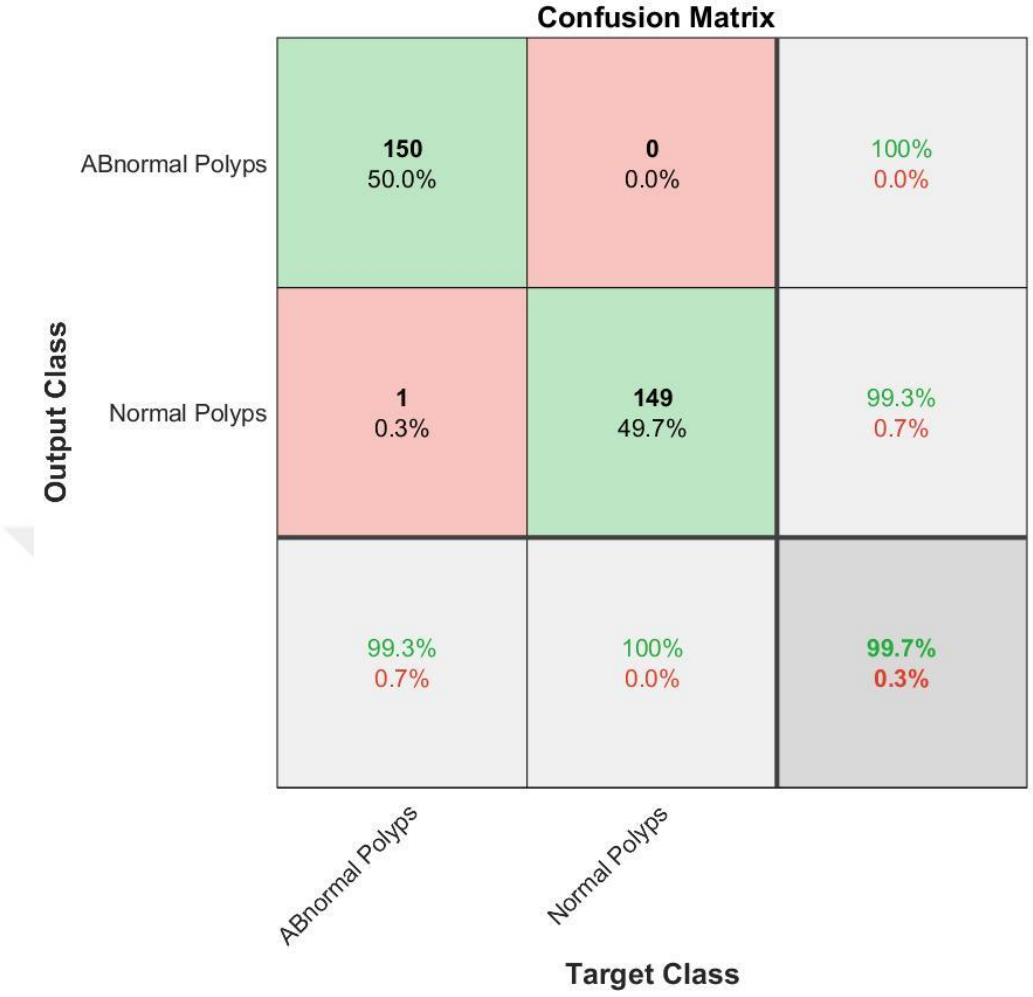

Figure 4.36 Our Method Results K1 Using SoftMax Classifier With One CNN Layer

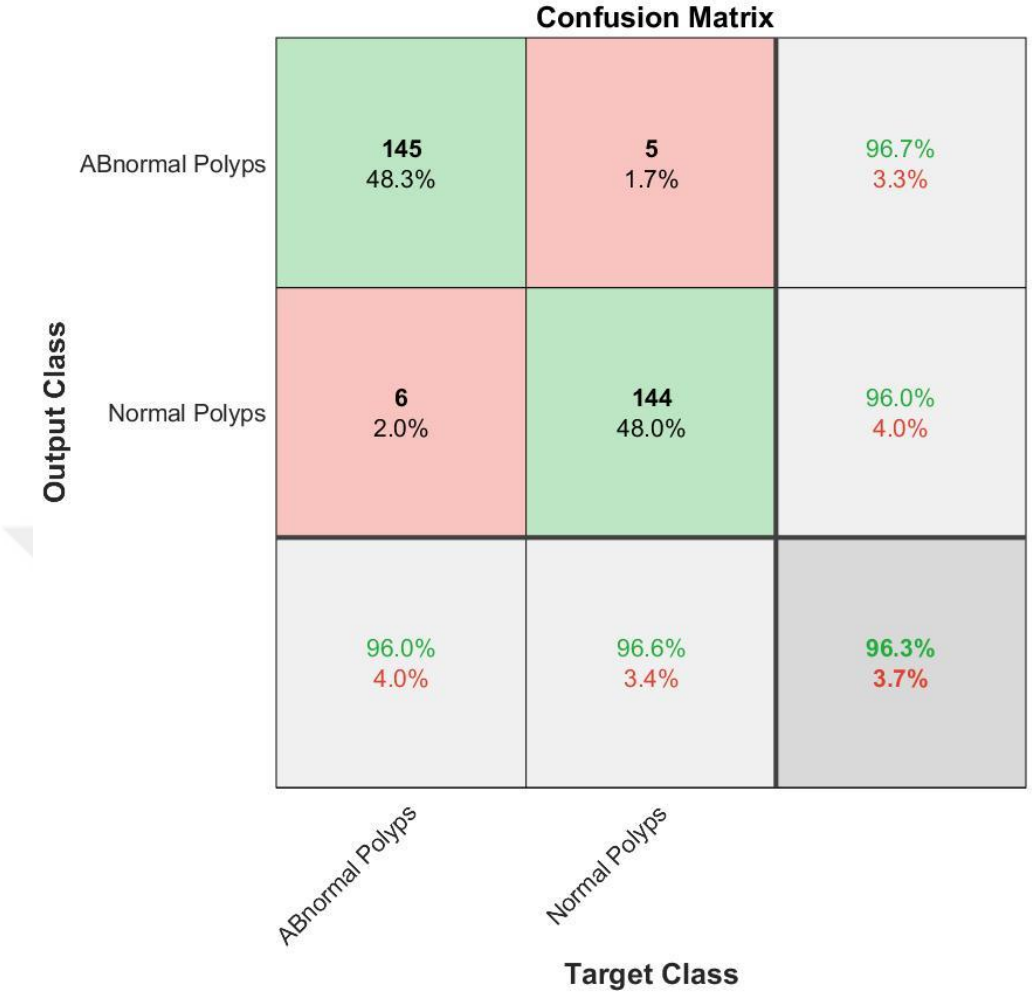

Figure 4.37 Our Method Results K2 Using SoftMax Classifier With One CNN Layer

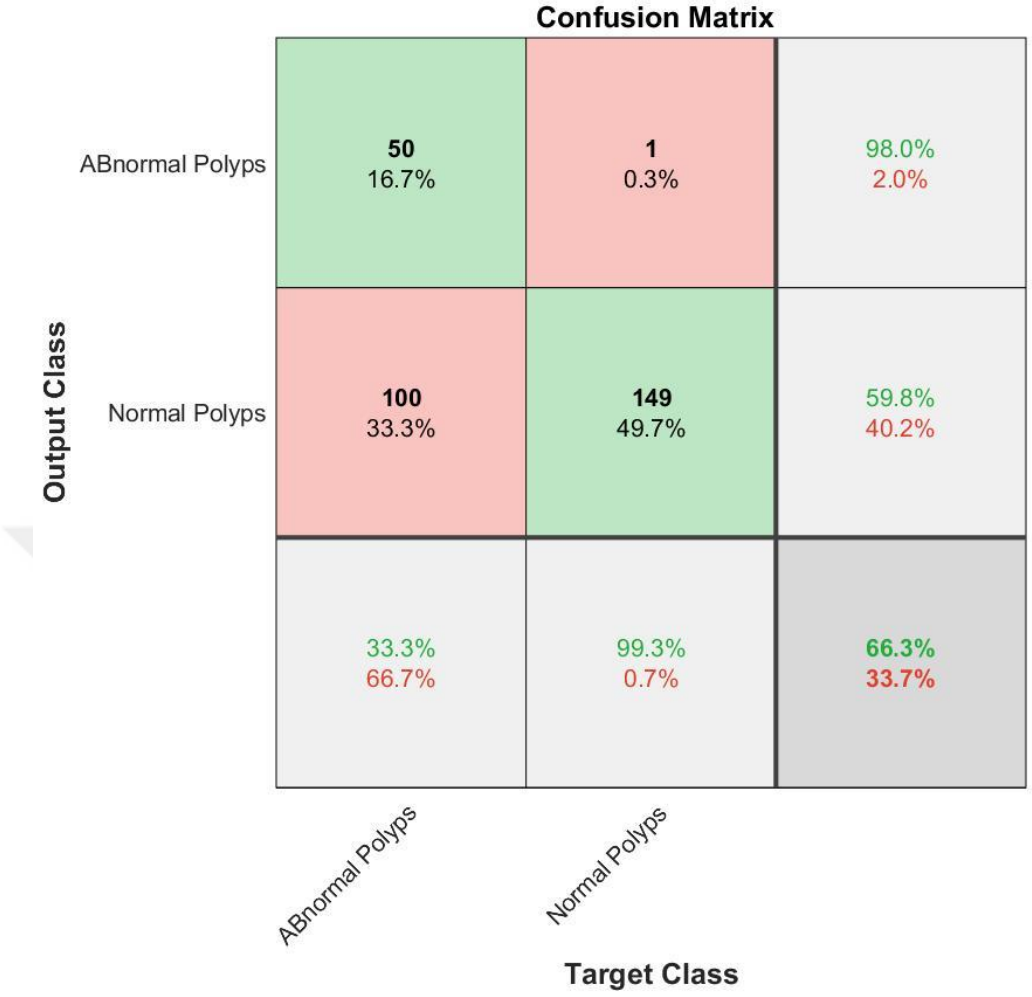

Figure 4.38 Our Method Results K3 Using SoftMax Classifier With One CNN Layer

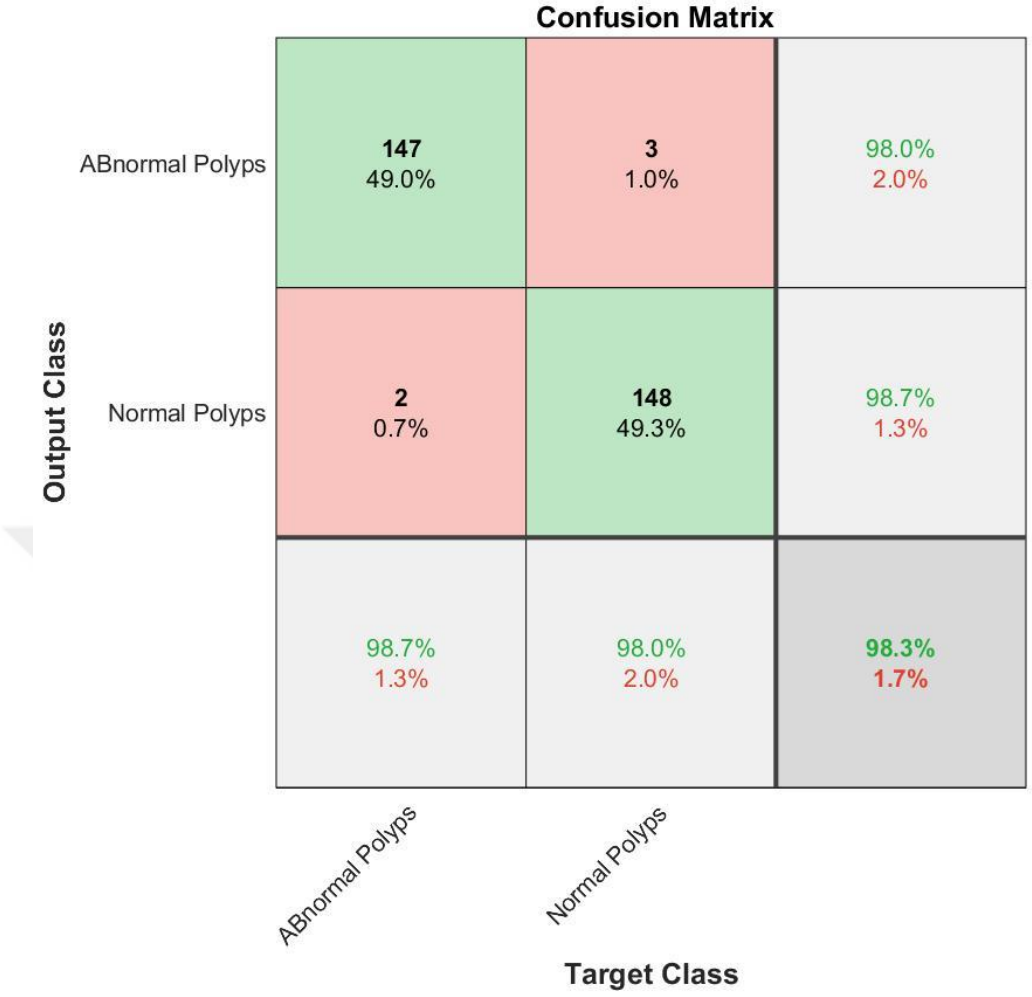

Figure 4.39 Our Method Results K4 Using SoftMax Classifier With One CNN Layer

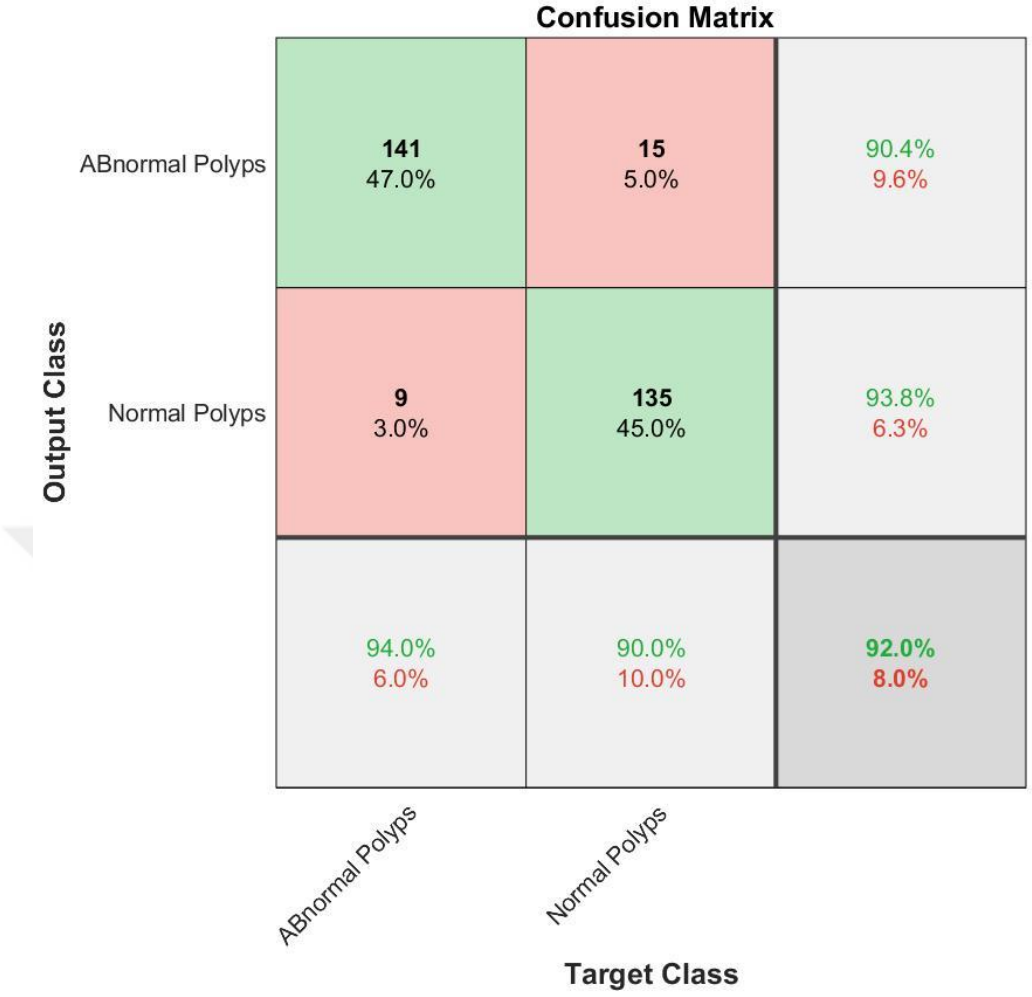

Figure 4.40 Our Method Results K5 Using SoftMax Classifier With One CNN Layer

Table 4.3 One CNN Layer Results

K-Fold	Test Results (%)	Time for Execution (Min)
K1	99.7	2:02
K2	96.3	1:56
K3	66.7	1:56
K4	98.3	1:58
K5	92.0	1:55
Average =	**90.6**	**1:57**

Now, choosing two layers of CNN with SoftMax classifier directly with five K-Fold cross validation are executed and the average of it is presented in bold in the table 4.4

and these results for each K are represneted in figure 4.41, 4.42, 4.43, 4.44 and figure 4.45.

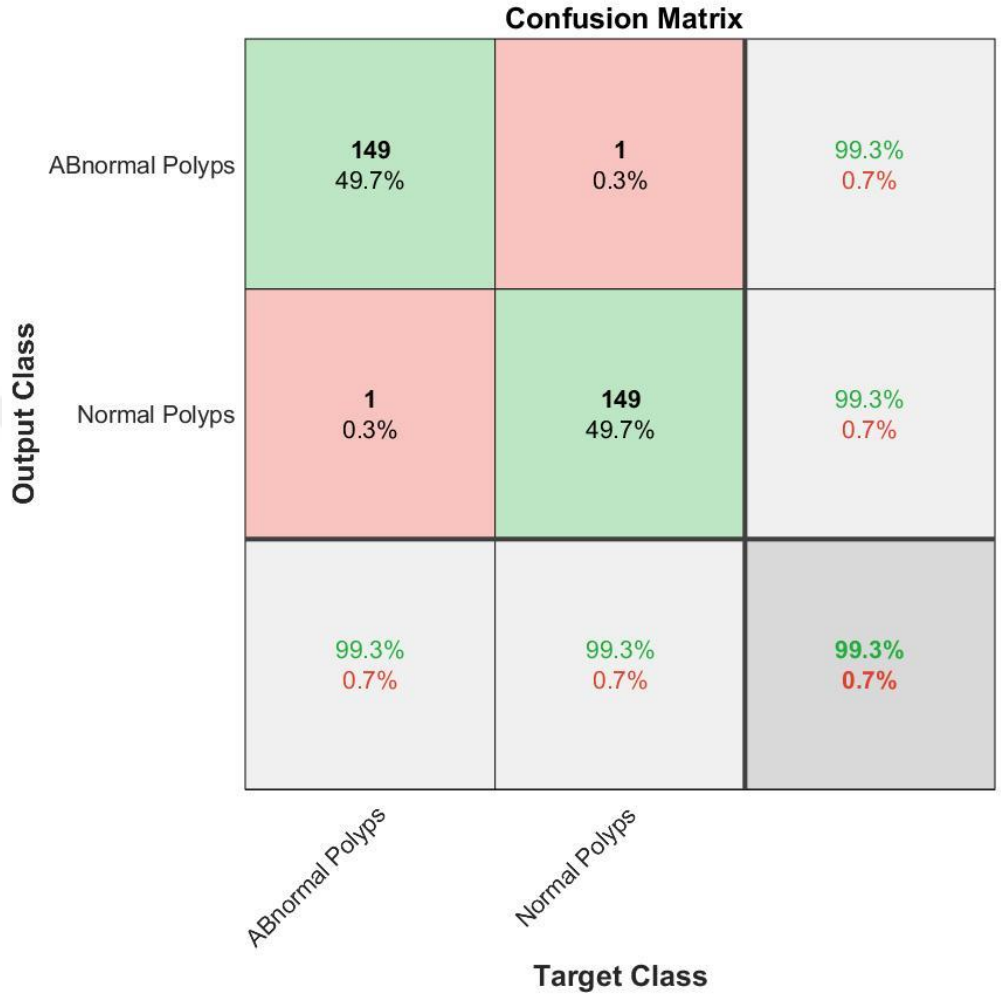

Figure 4.41 Our Method Results K1 Using SoftMax Classifier With Two CNN Layers

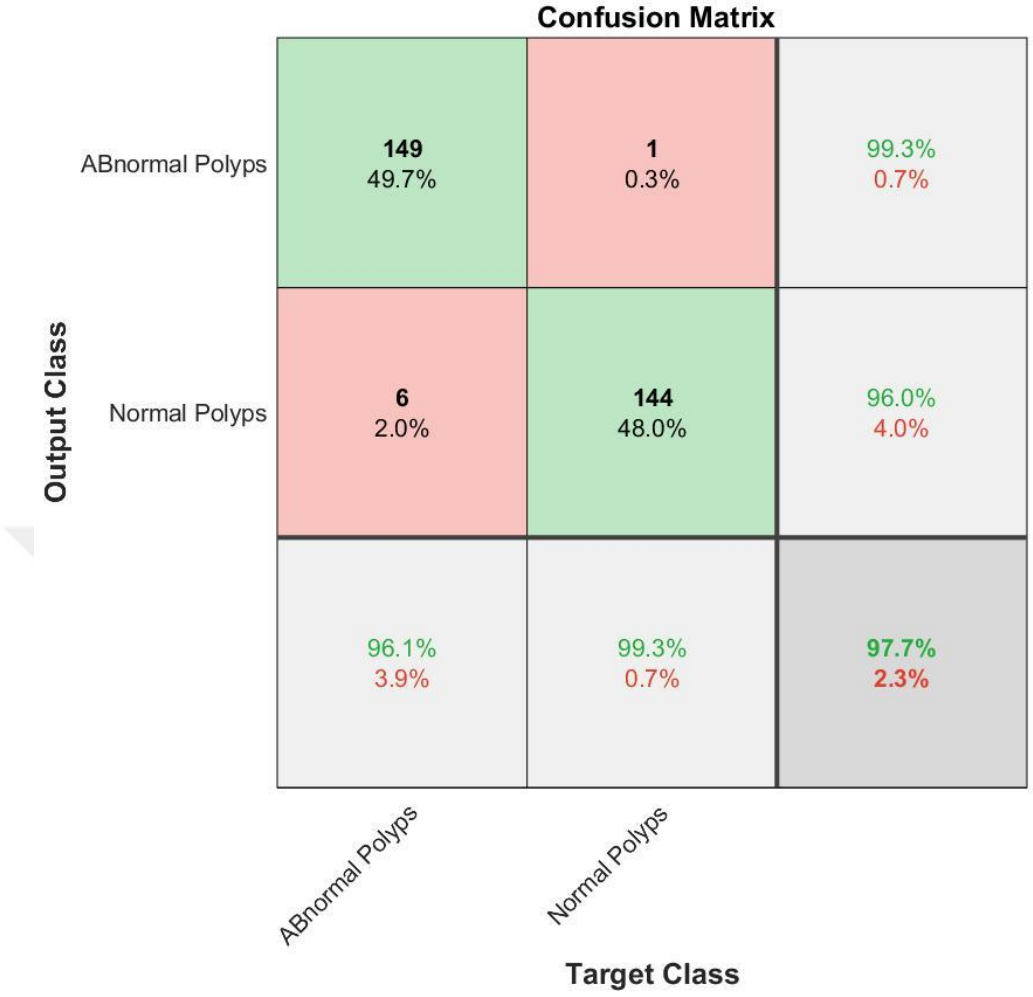

Figure 4.42 Our Method Results K2 Using SoftMax Classifier With Two CNN Layers

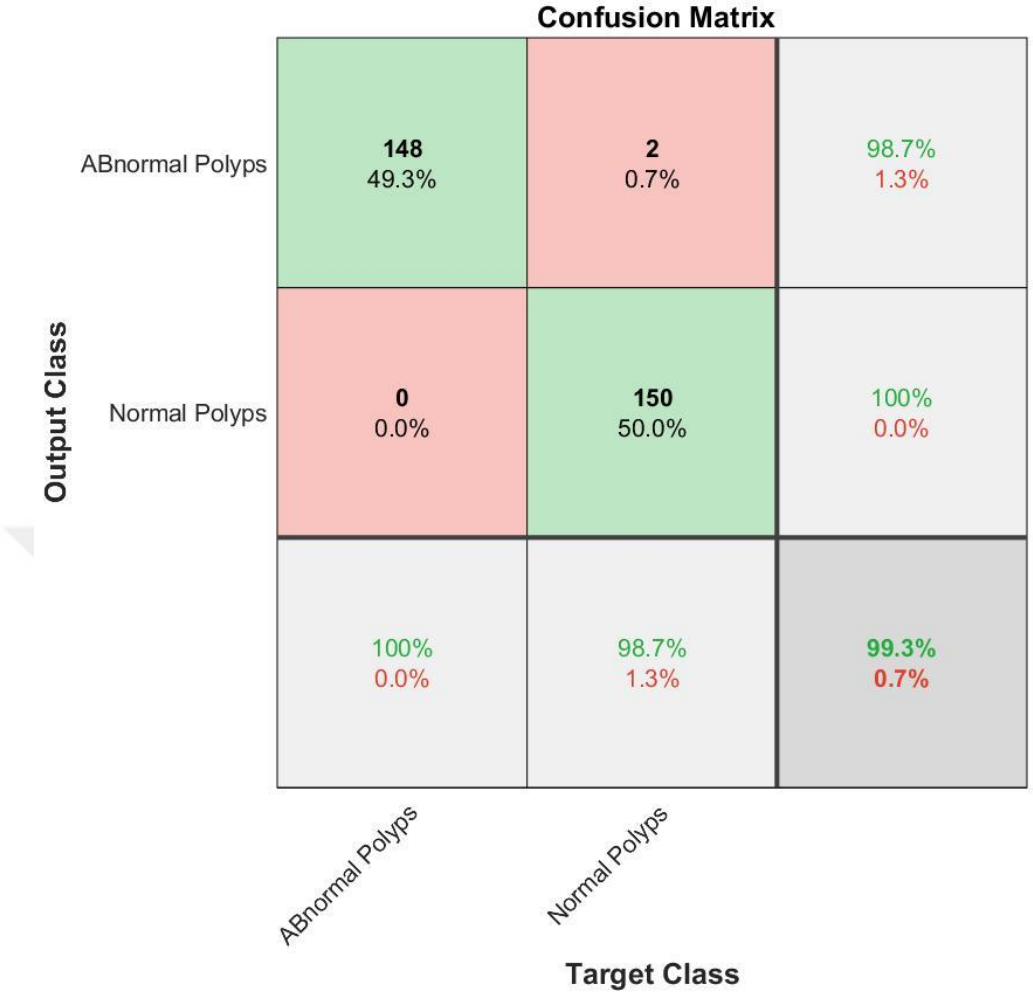

Figure 4.43 Our Method Results K3 Using SoftMax Classifier With Two CNN Layers

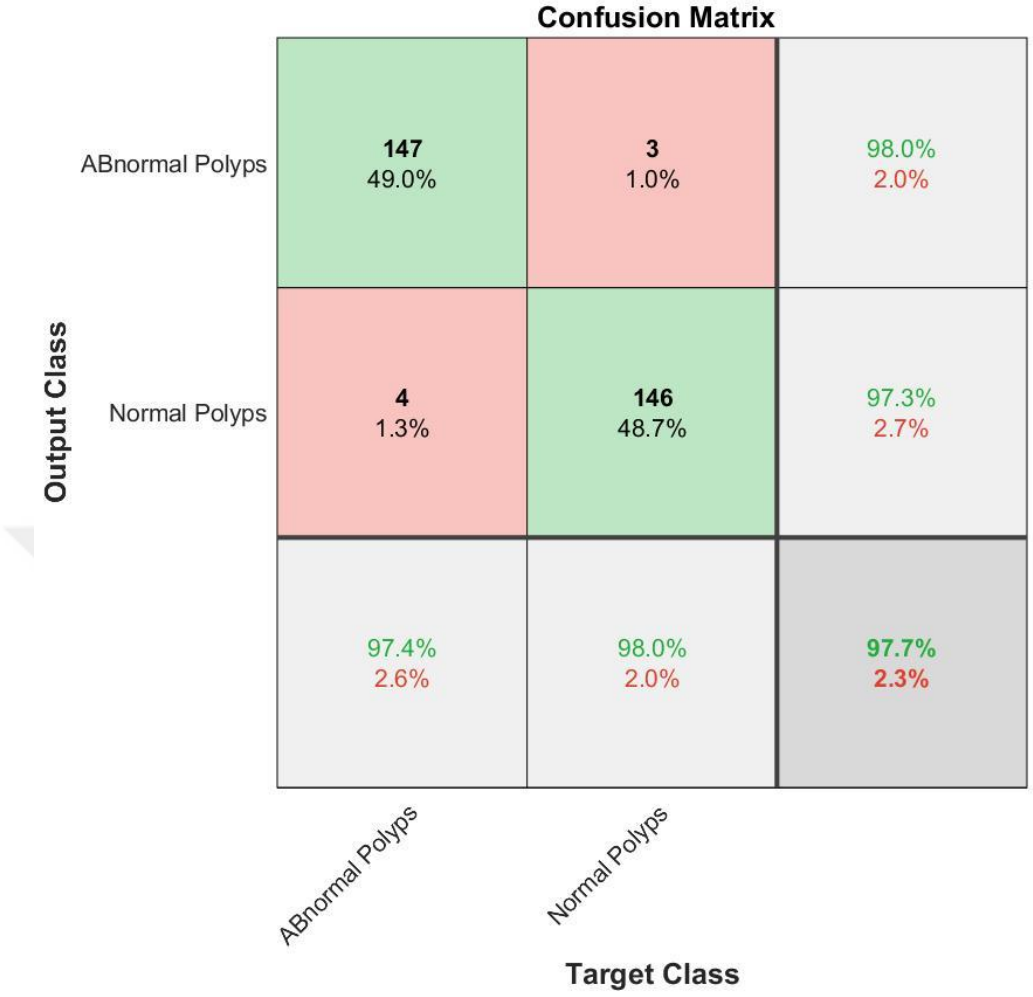

Figure 4.44 Our Method Results K4 Using SoftMax Classifier With Two CNN Layers

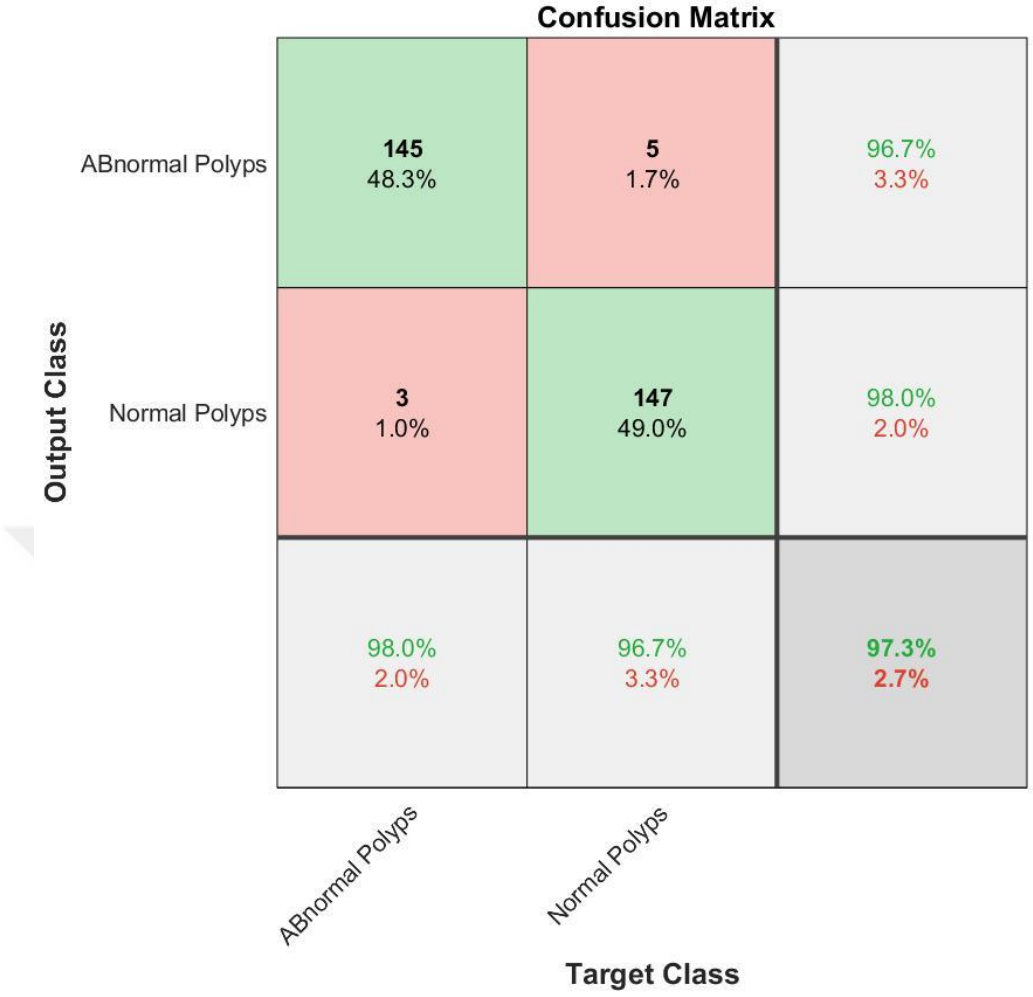

Figure 4.45 Our Method Results K5 Using SoftMax Classifier With Two CNN Layers

Table 4.4 Two CNN Layers Results

K-Fold	Test Results (%)	Time for Execution (Min)
K1	99.3	2:49
K2	97.7	2:34
K3	99.3	2:30
K4	97.7	2:31
K5	97.3	2:34
Average =	**98.26**	**2:35**

instead of two layers of CNN we prefer to choose three CNN layers with same classifier and same properties, SoftMax classifier and applying K-Fold cross validation with five fold which these are the best select for our problem it is bring an average 99.88% for classification and acceptable time 3:05 min. presented in bold in the table 4.5 and these results for each K are represneted in figure 4.46, 4.47, 4.48, 4.49 and figure 4.50.

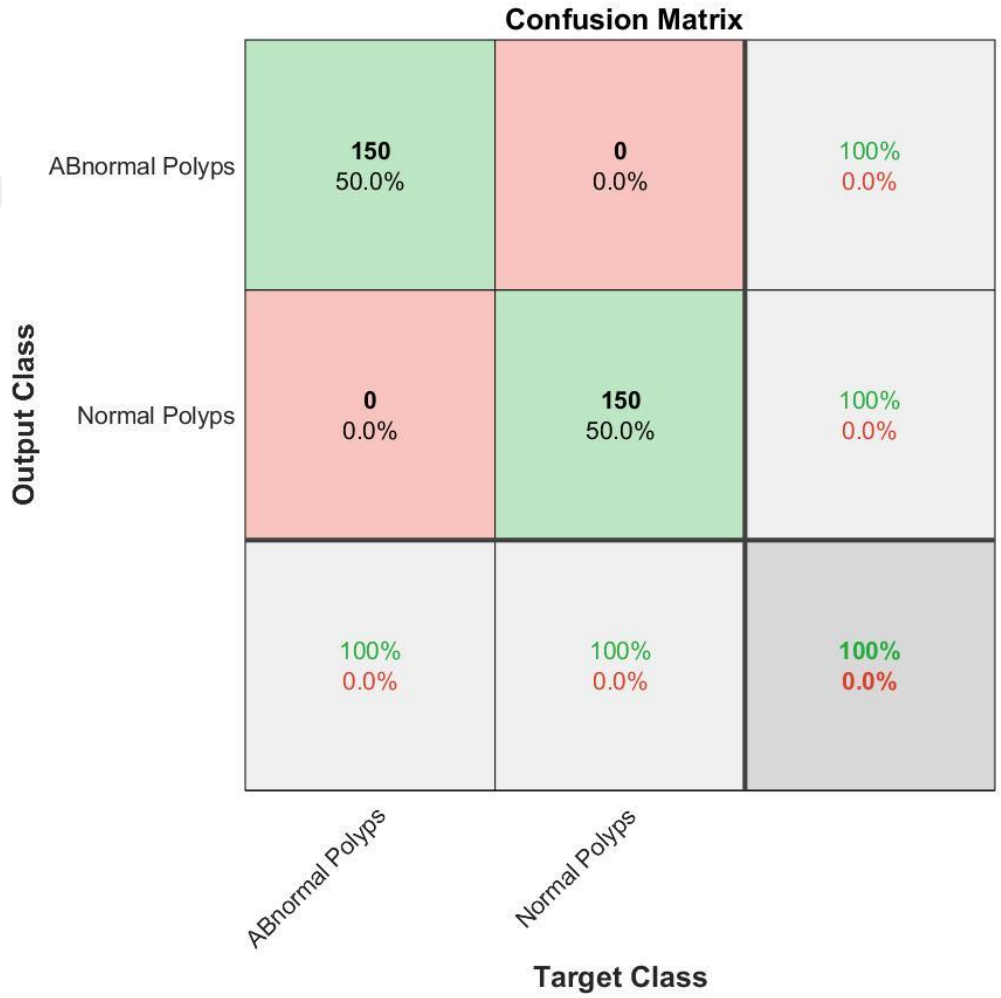

Figure 4.46 Our Method Results K1 Using SoftMax Classifier With Three CNN Layers

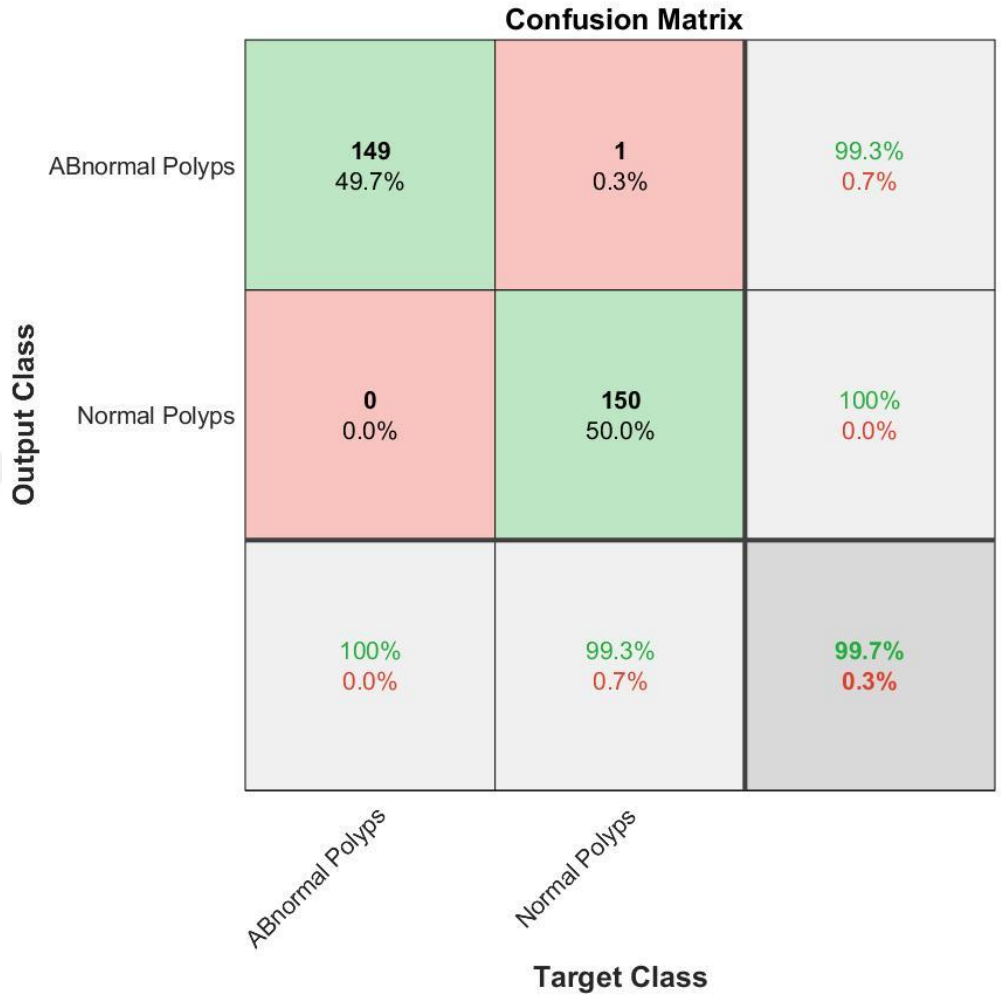

Figure 4.47 Our Method Results K2 Using SoftMax Classifier With Three CNN Layers

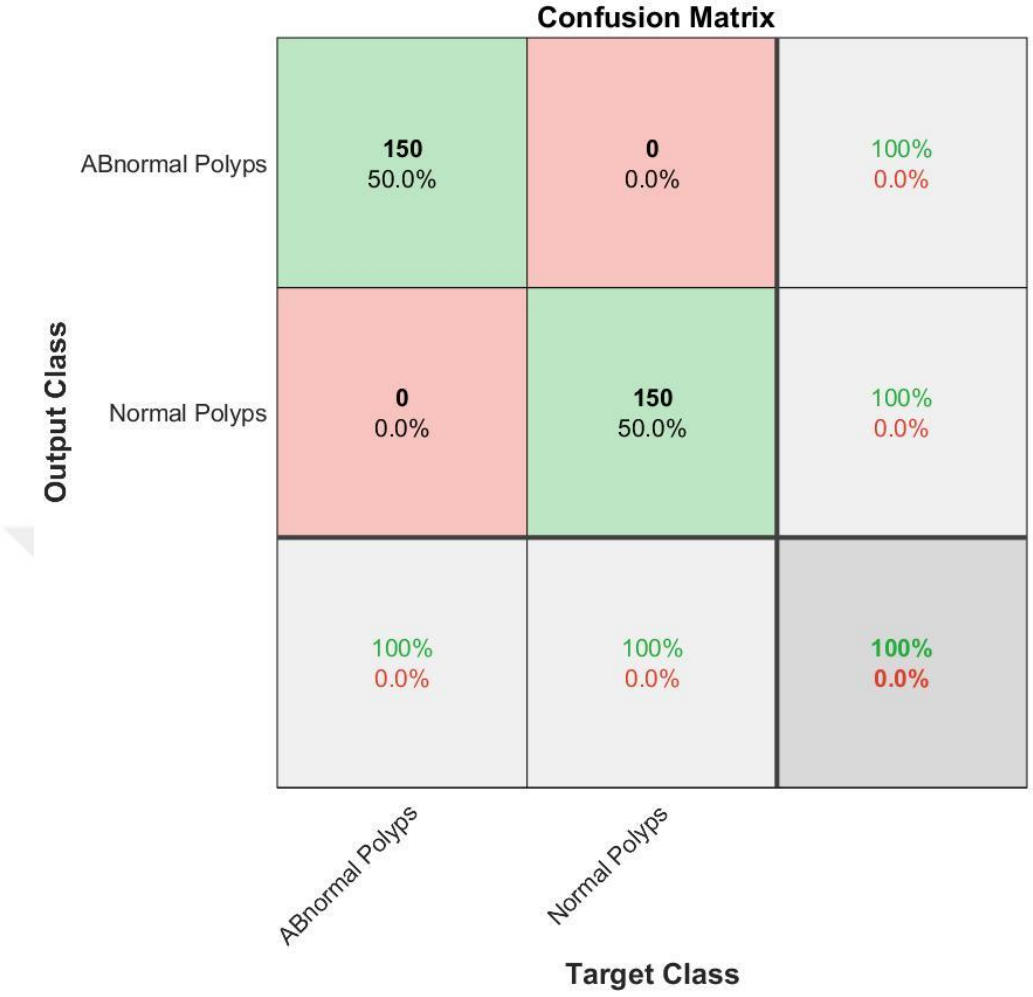

Figure 4.48 Our Method Results K3 Using SoftMax Classifier With Three CNN Layers

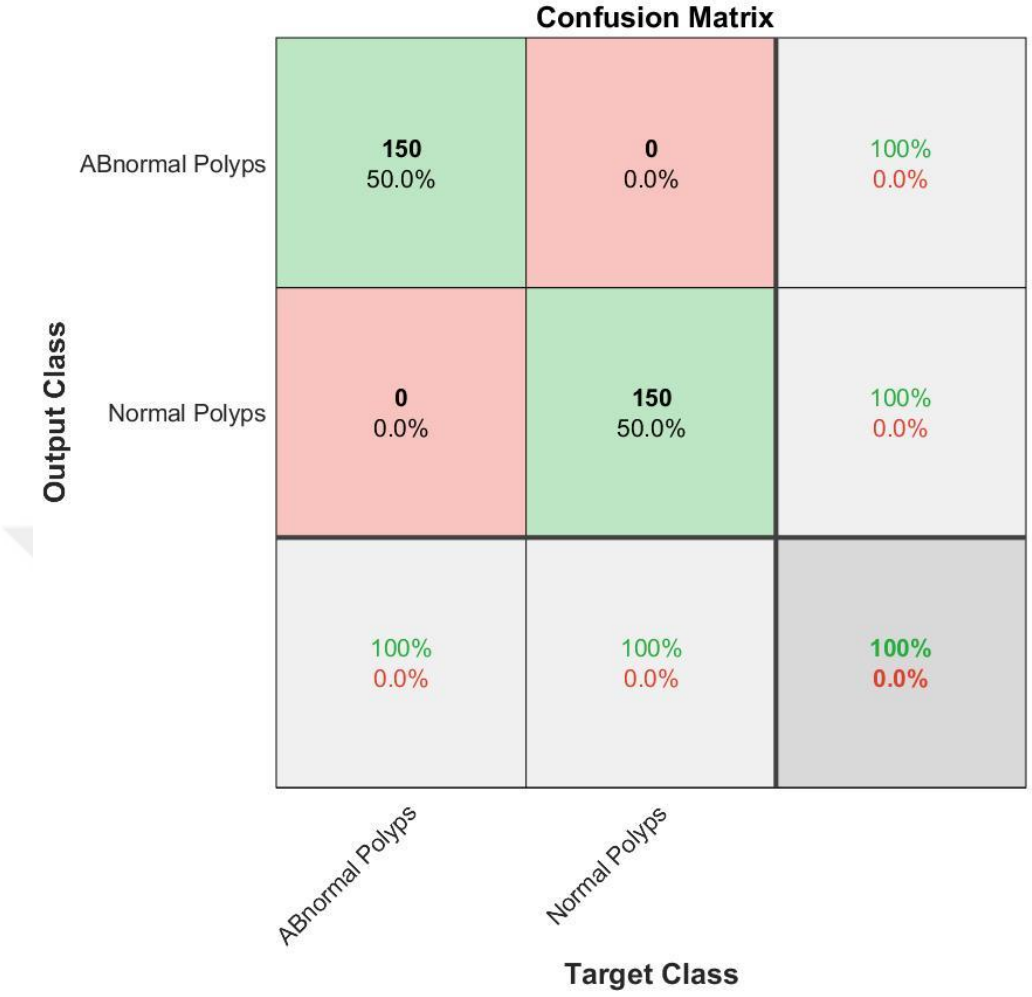

Figure 4.49 Our Method Results K4 Using SoftMax Classifier With Three CNN Layers

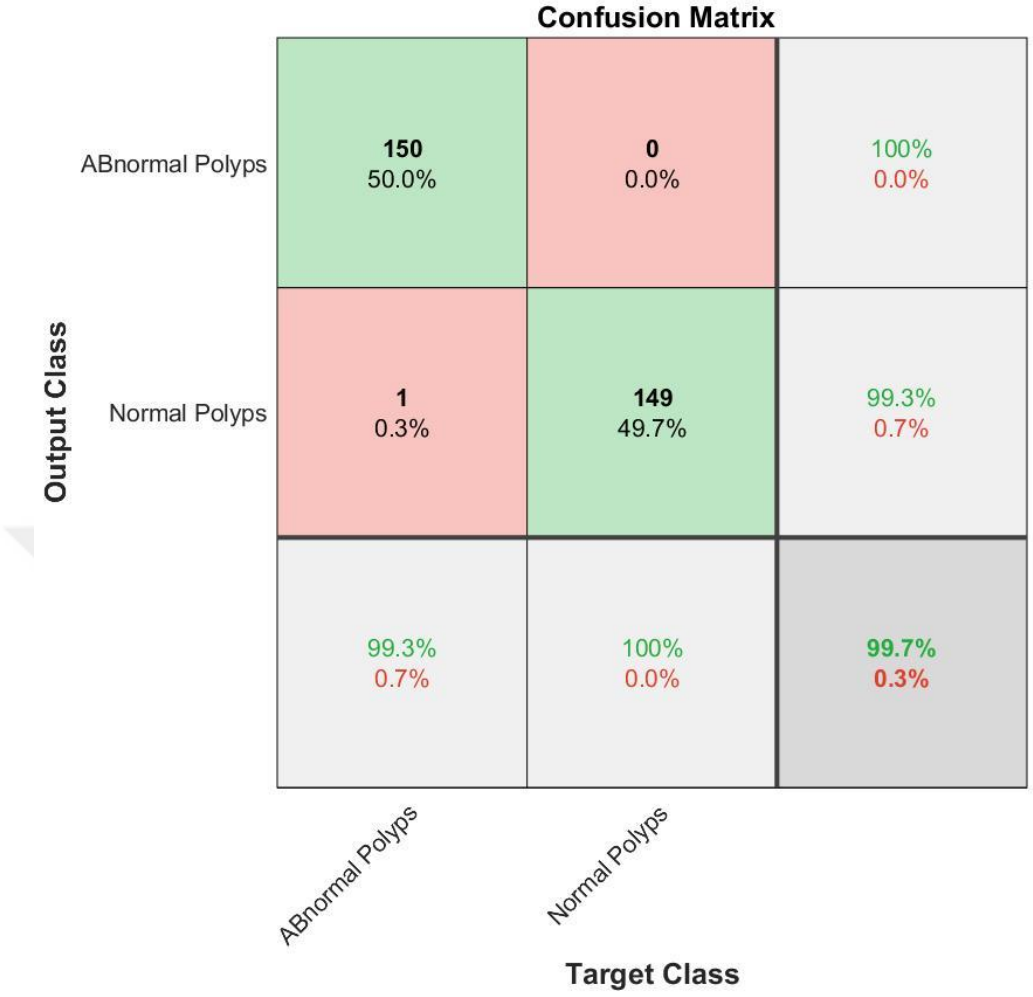

Figure 4.50 Our Method Results K5 Using SoftMax Classifier With Three CNN Layers

Table 4.5 Three CNN Layers Results

K-Fold	Test Results (%)	Time for Execution (Min)
K1	100	3:15
K2	99.7	3:05
K3	100	3:01
K4	100	3:01
K5	99.7	3:02
Average =	**99.88**	**3:05**

Also four CNN layers applied to the same problem, which has less accuracy and more time to implement the work comparing with thee CNN layers, and these results shows in table 4.5 which includes five fold cross validation and average for them, these results for each K are represneted in figure 4.51, 4.52, 4.53, 4.54 and figure 4.55.

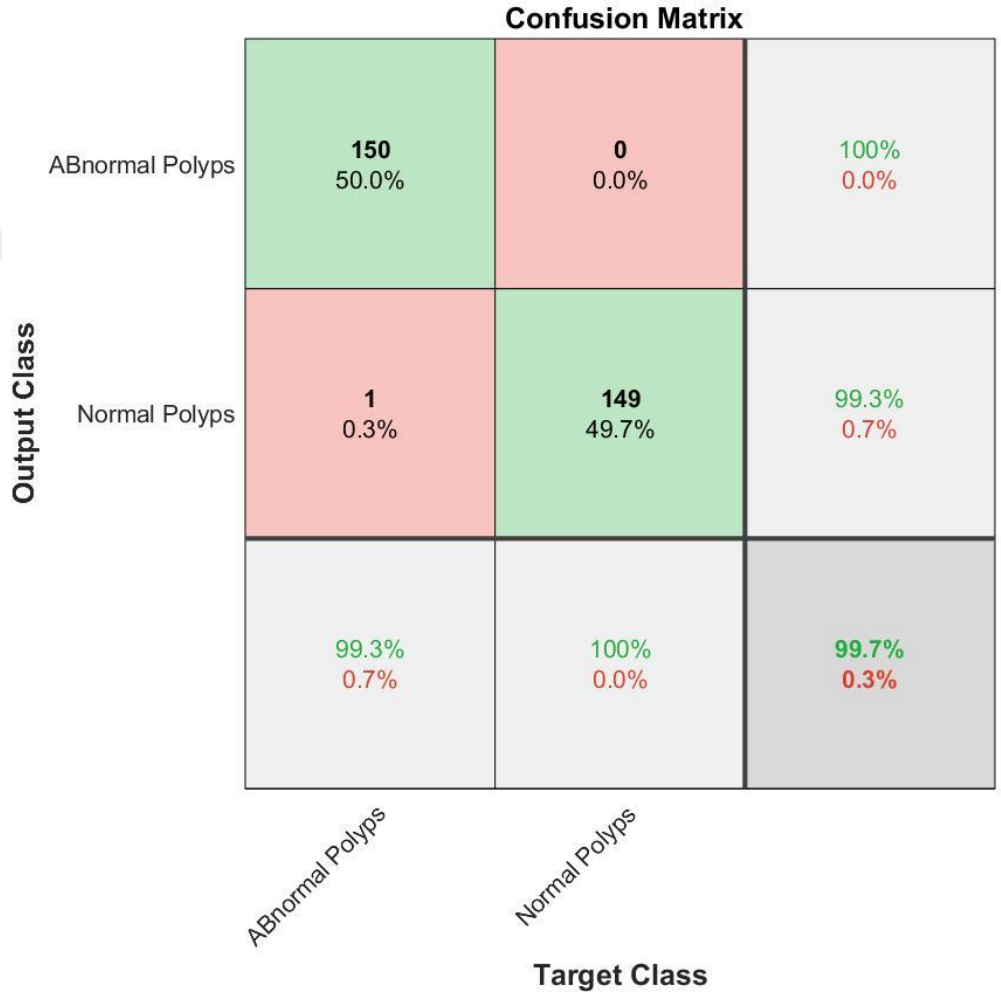

Figure 4.51 Our Method Results K1 Using SoftMax Classifier With Four CNN Layers

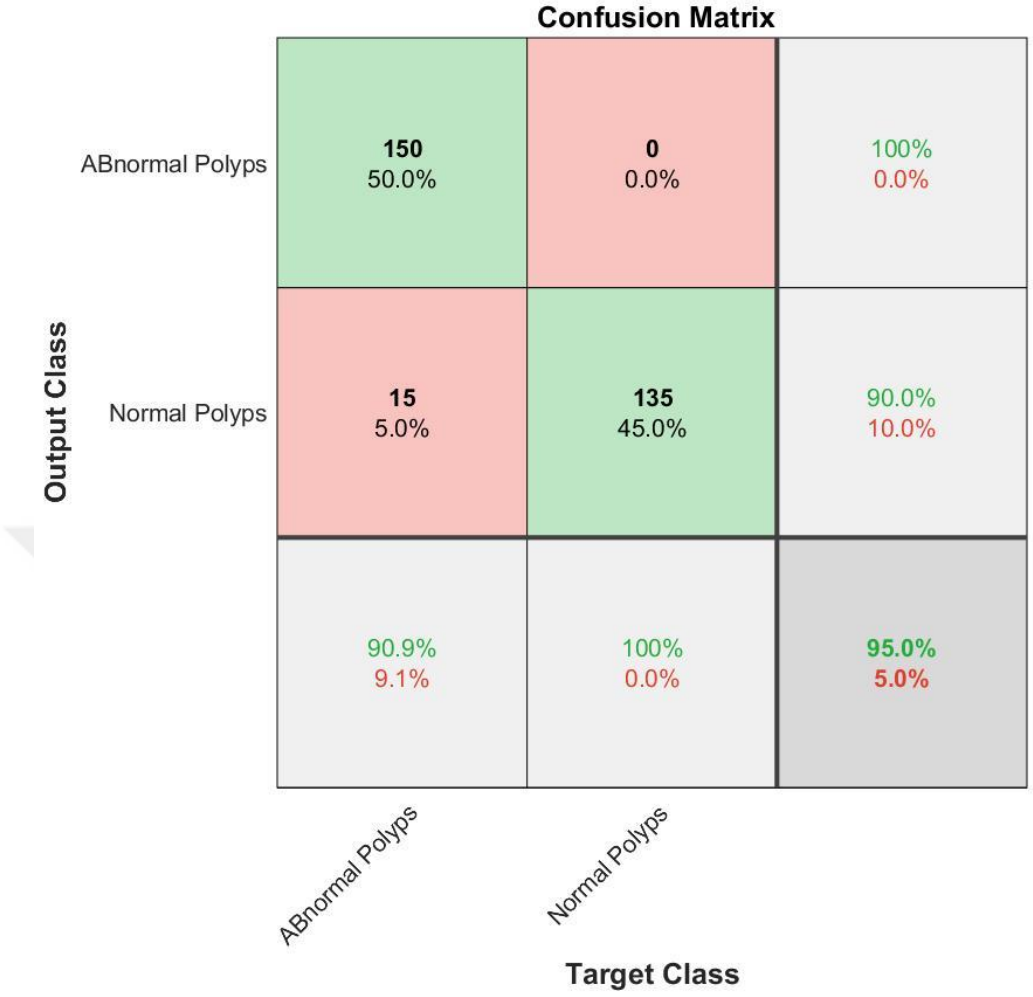

Figure 4.52 Our Method Results K2 Using SoftMax Classifier With Four CNN Layers

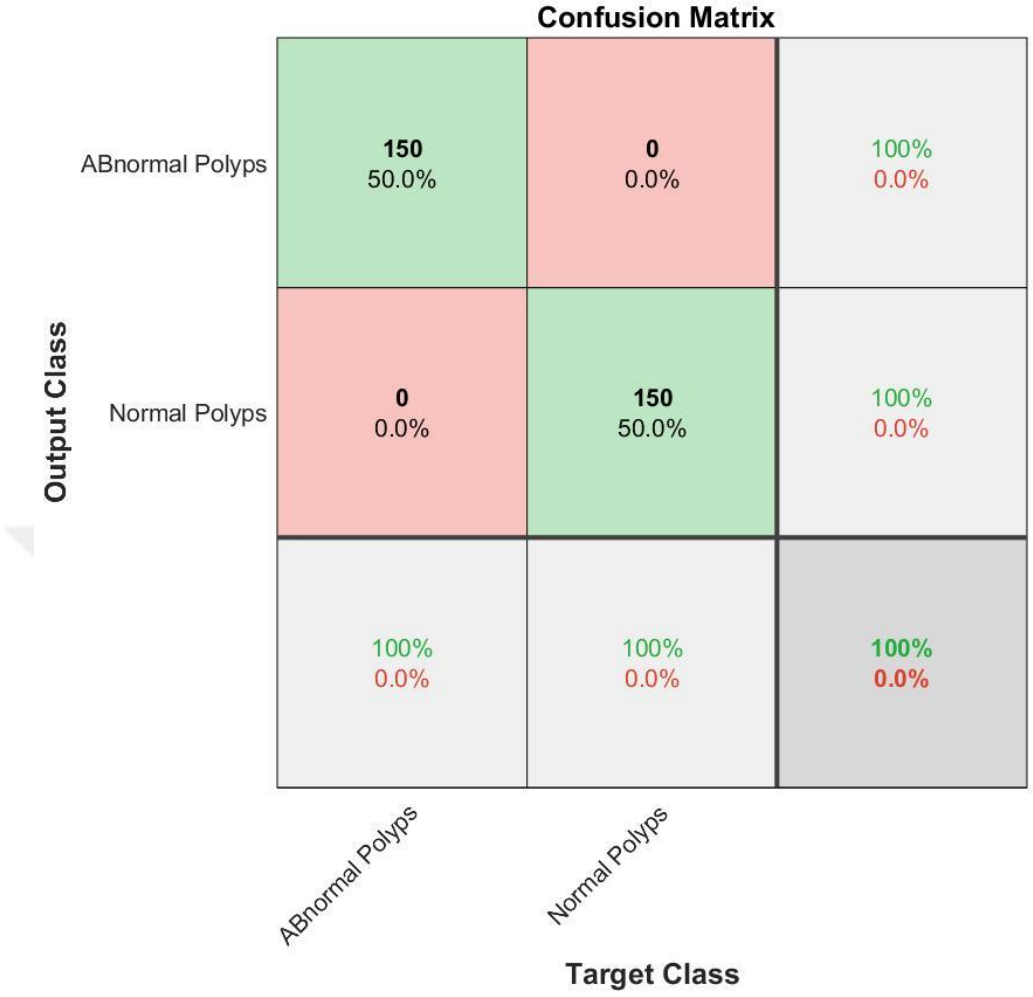

Figure 4.53 Our Method Results K3 Using SoftMax Classifier With Four CNN Layers

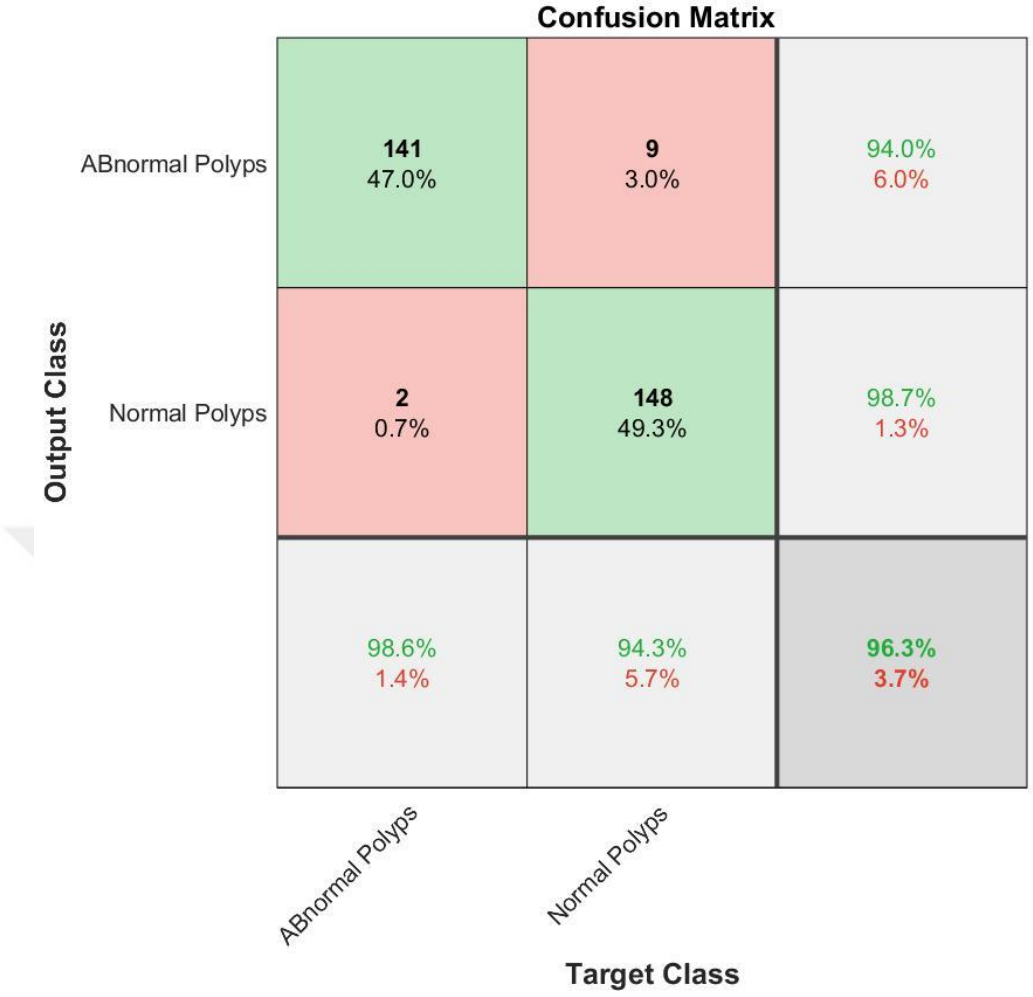

Figure 4.54 Our Method Results K4 Using SoftMax Classifier With Four CNN Layers

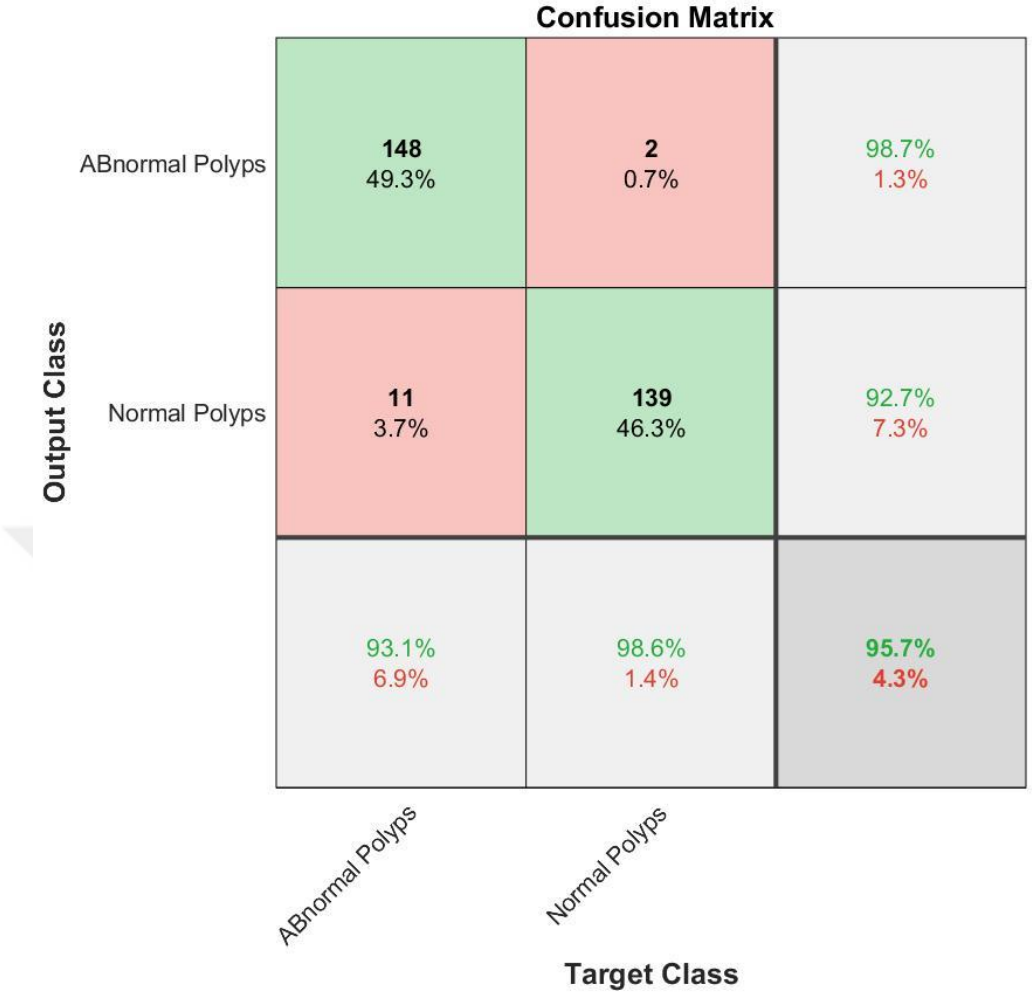

Figure 4.55 Our Method Results K5 Using SoftMax Classifier With Four CNN Layers

Table 4.6 Four CNN Layers Results

K-Fold	Test Results (%)	Time for Execution (Min)
K1	99.7	8:00
K2	95.0	8:22
K3	100	9:06
K4	96.3	8:42
K5	95.7	9:19
Average =	**97.34**	**8:42**

Also five CNN layers applied to the same problem, which it present it take long time for execution with less accuracy and it is approximately stable results comparing with other number of layers except the three CNN layers which is the best choose, average result for five CNN layers with five fold cross validation and the average of these obtained in bold in table 4.7 and these results for each K are represneted in figure 4.56, 4.57, 4.58, 4.59 and figure 4.60.

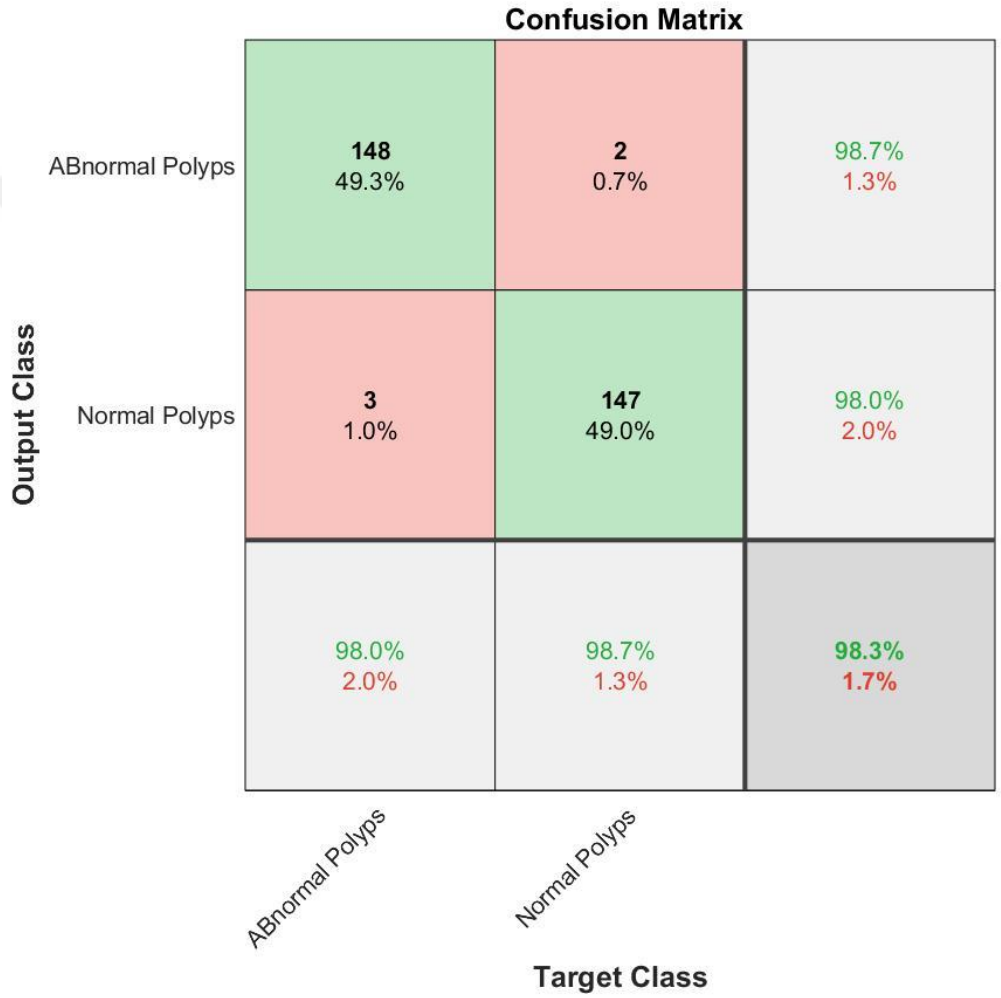

Figure 4.56 Our Method Results K1 Using SoftMax Classifier With Five CNN Layers

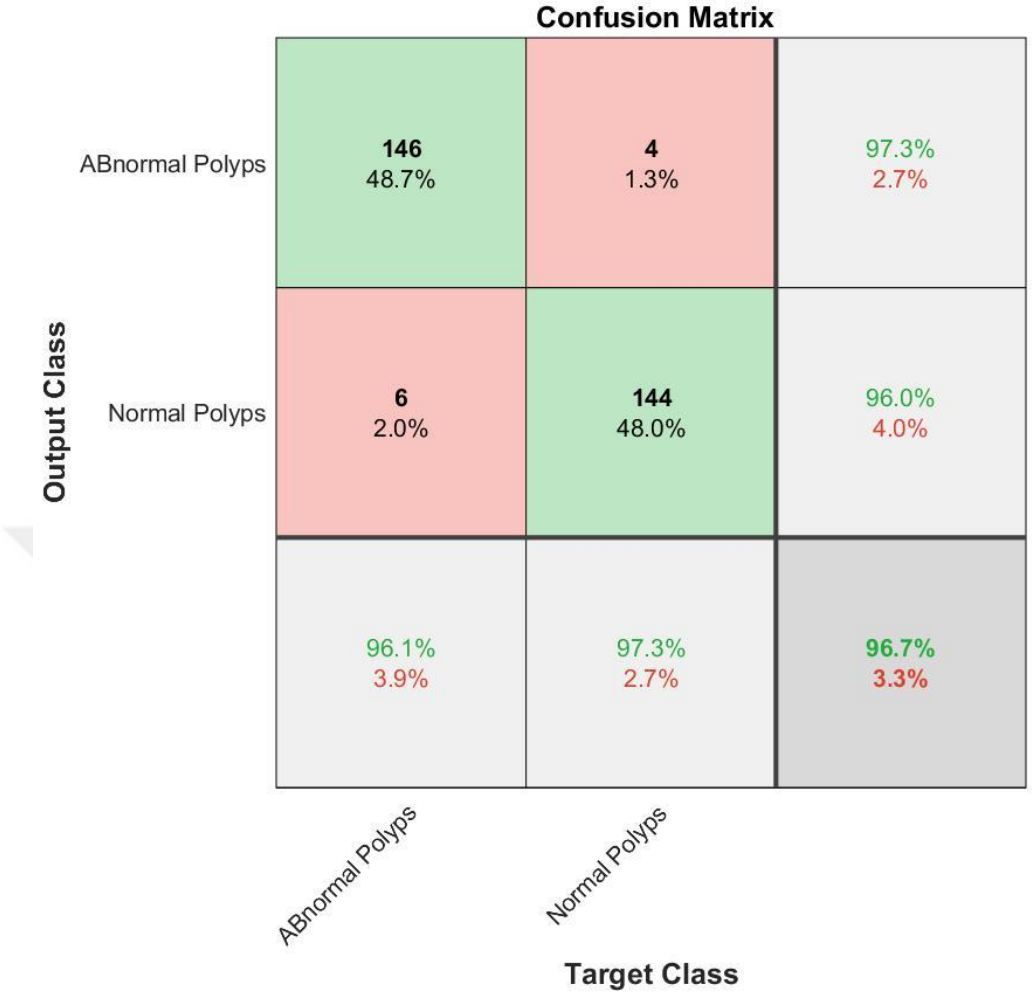

Figure 4.57 Our Method Results K2 Using SoftMax Classifier With Five CNN Layers

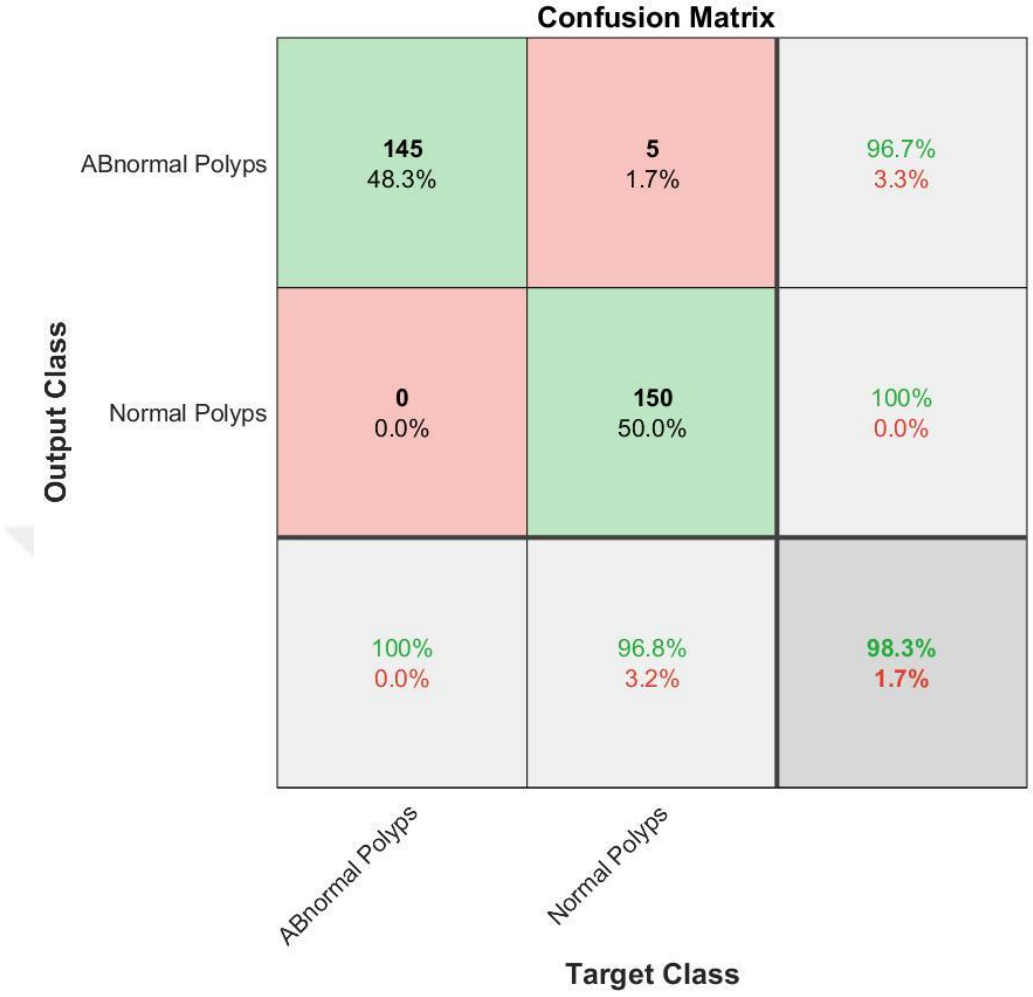

Figure 4.58 Our Method Results K3 Using SoftMax Classifier With Five CNN Layers

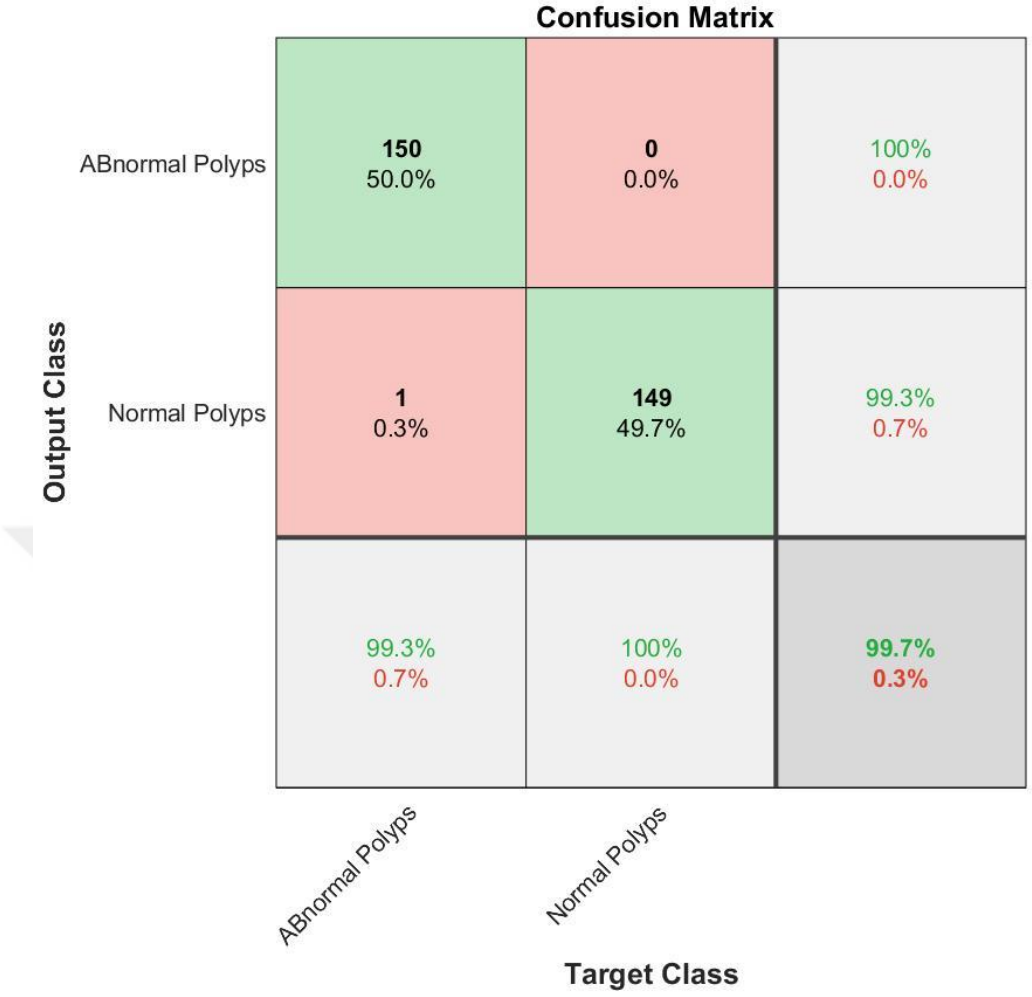

Figure 4.59 Our Method Results K4 Using SoftMax Classifier With Five CNN Layers

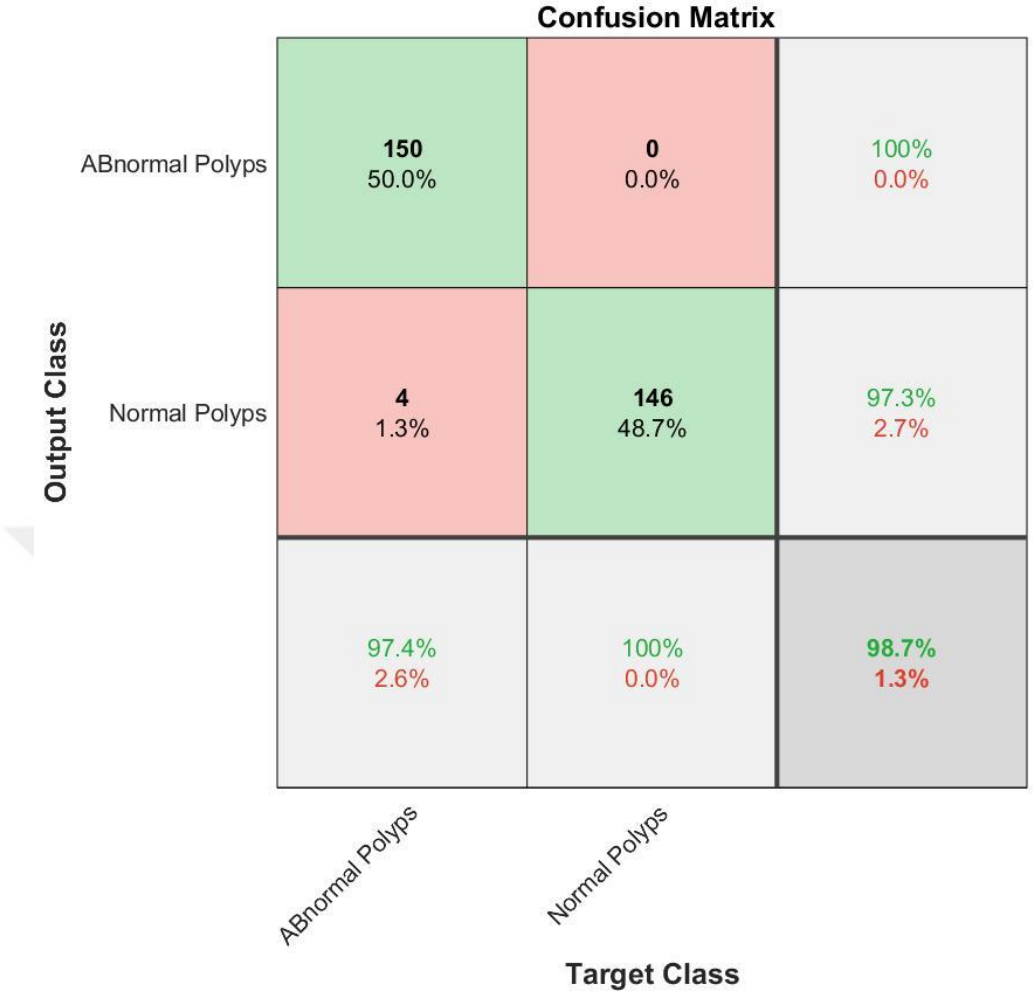

Figure 4.60 Our Method Results K5 Using SoftMax Classifier With Five CNN Layers

Table 4.7 Five CNN Layers Results

K-Fold	Test Results (%)	Time for Execution (Min)
K1	98.3	9:27
K2	96.7	9:51
K3	98.3	8:31
K4	99.7	8:43
K5	98.7	8:45
Average =	**98.34**	**9:03**

Also six layers applied to the same problem, which mean using six layers to solve detection problems. The experiments results presented in the figure 4.61, 4.62, 4.63, 4.64, and figure 4.65. The Table 4.8 show the average accuracy of the experiments and execution time.

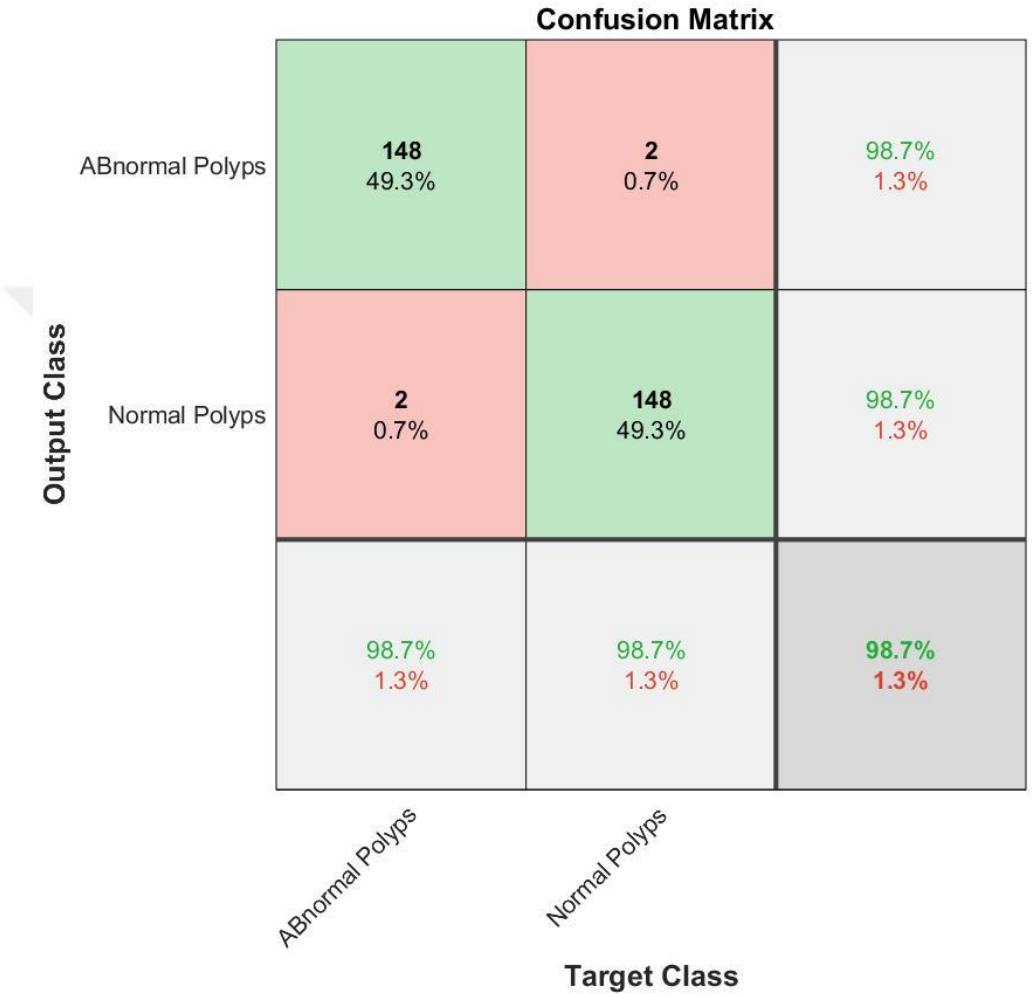

Figure 4.61 Our Method Results K1 Using SoftMax Classifier With Six CNN Layers

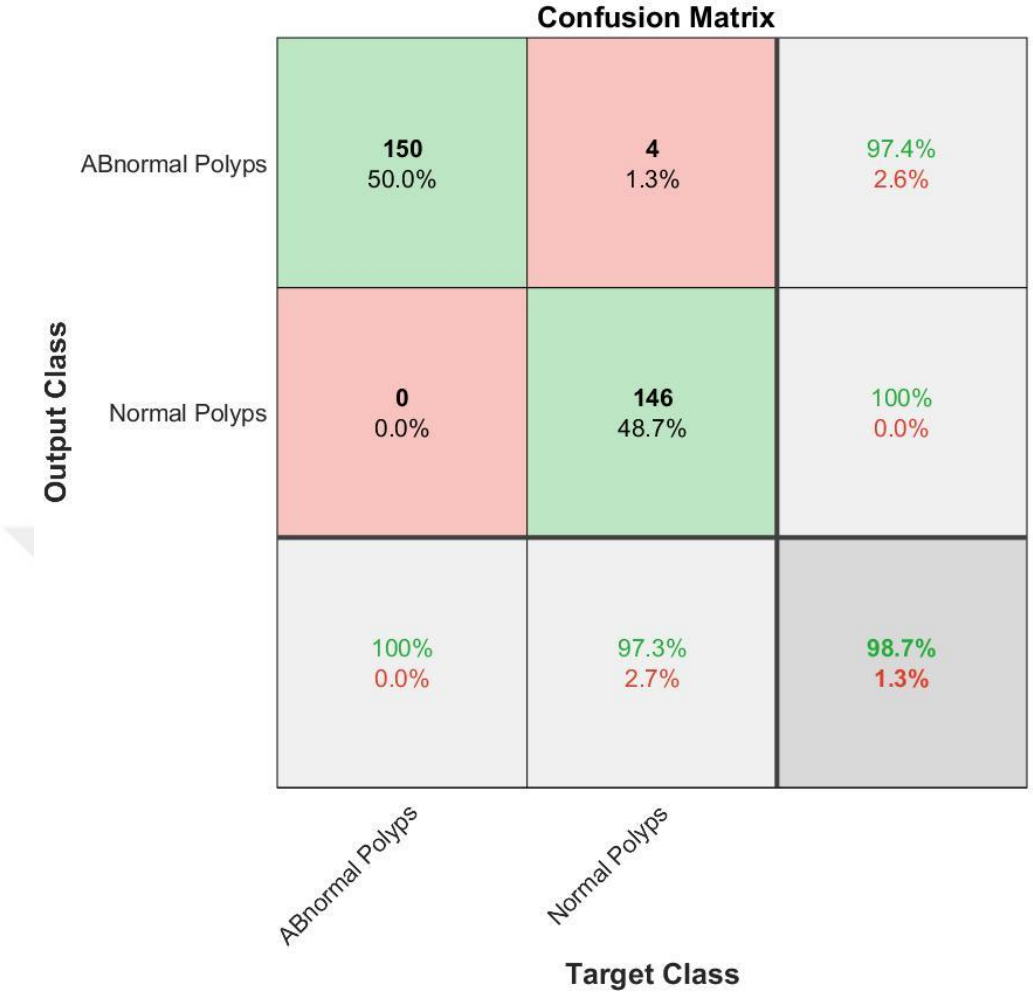

Figure 4.62 Our Method Results K2 Using SoftMax Classifier With Six CNN Layers

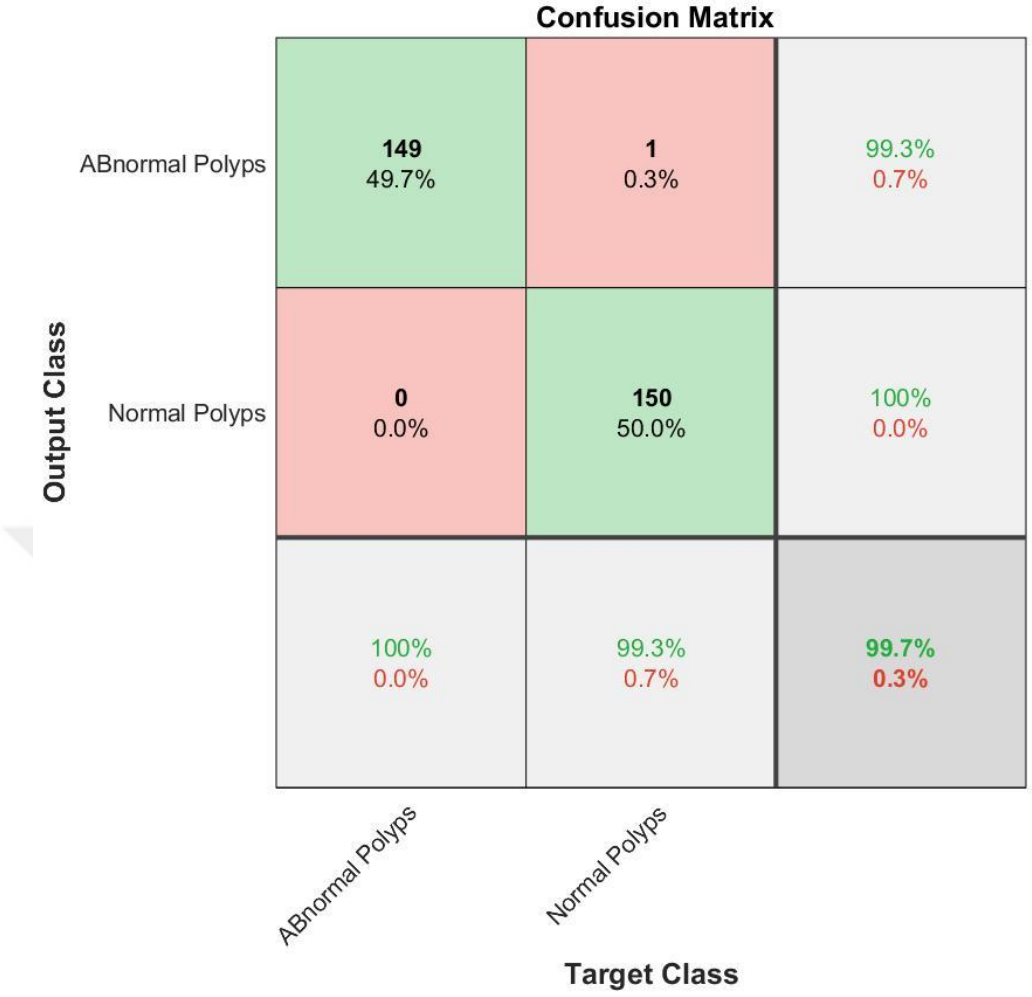

Figure 4.63 Our Method Results K3 Using SoftMax Classifier With Six CNN Layers

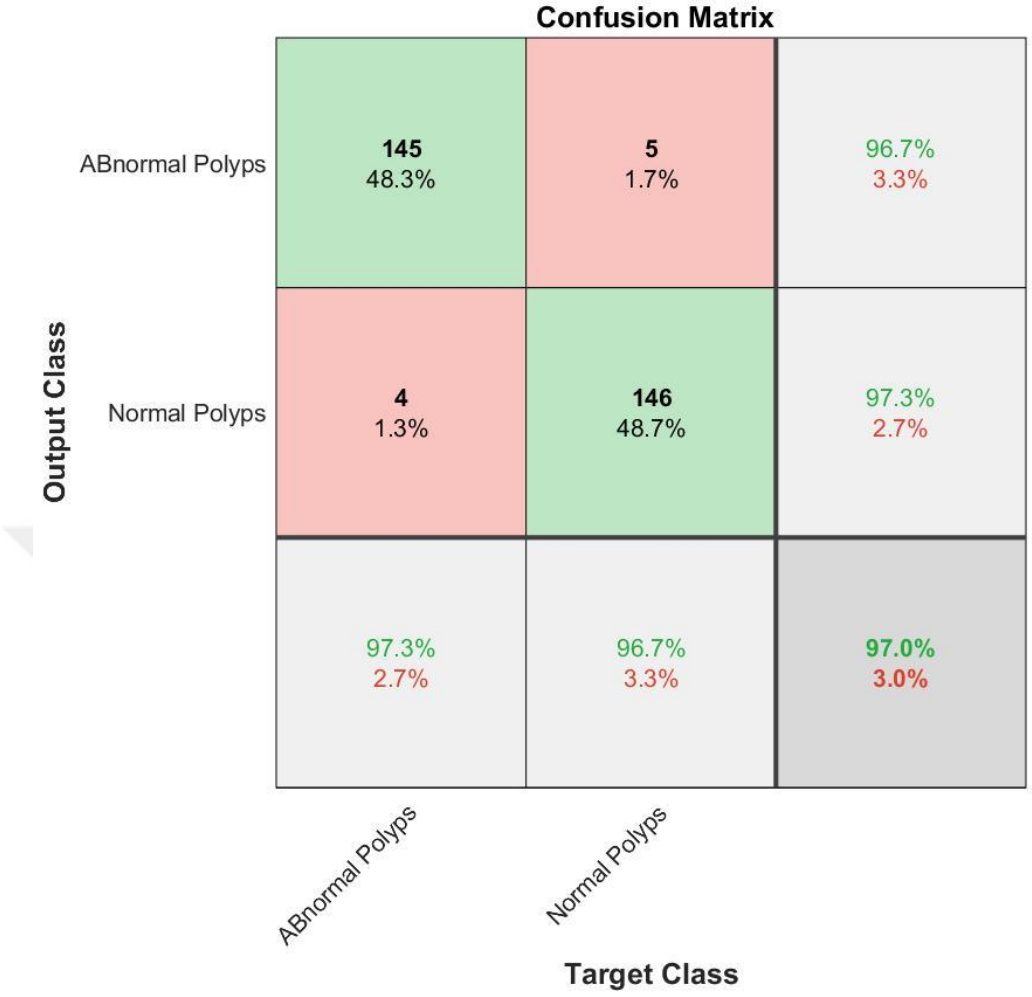

Figure 4.64 Our Method Results K4 Using SoftMax Classifier With Six CNN Layers

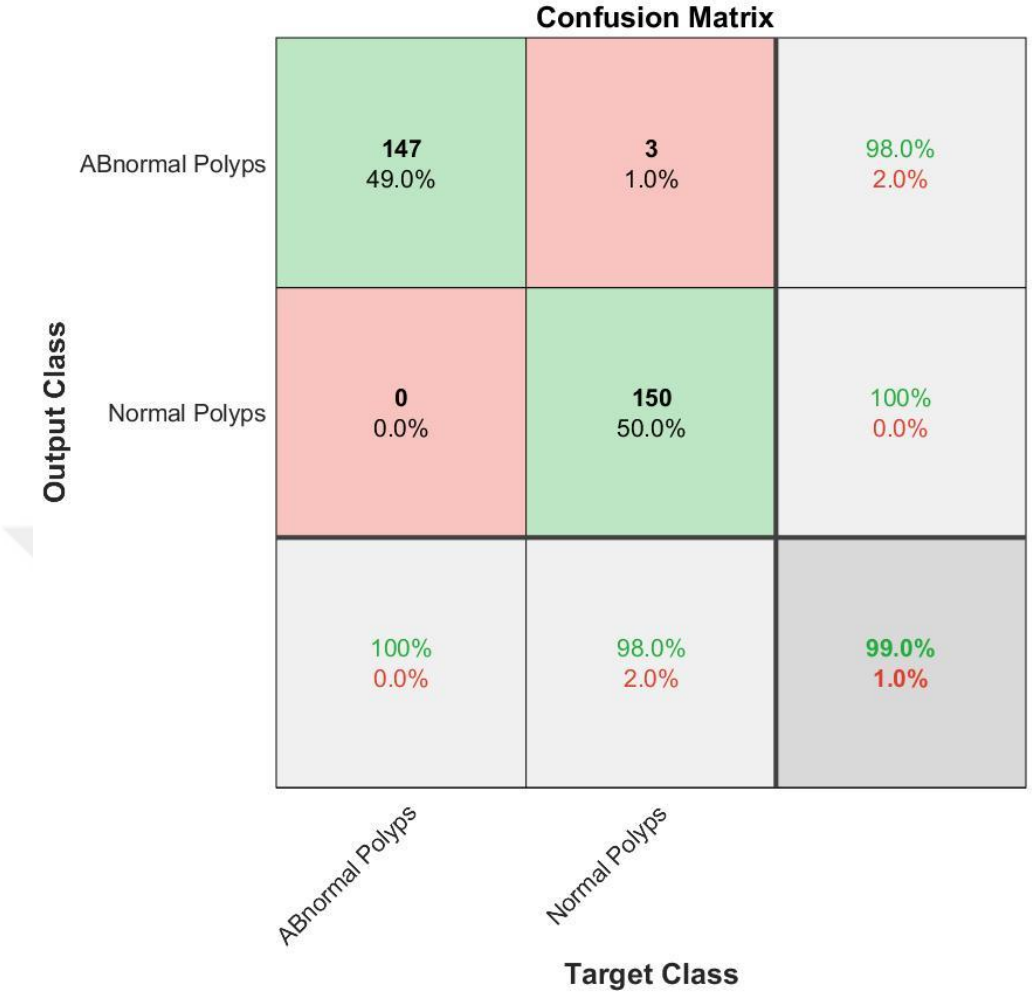

Figure 4.65 Our Method Results K5 Using SoftMax Classifier With Six CNN Layers

Table 4.8 Six CNN Layers Results

K-Fold	Test Results (%)	Time for Execution (Min)
K1	98.7	8:35
K2	98.7	10:51
K3	99.7	9:11
K4	97.0	10:56
K5	99.0	10:52
Average =	**98.62**	**10:05**

Also seven layers applied to the same problem. The experiments results presented in the figure 4.66, 4.67, 4.68, 4.69, and figure 4.70. The Table 4.9 presented the execution time and the avreage accuracy of the model.

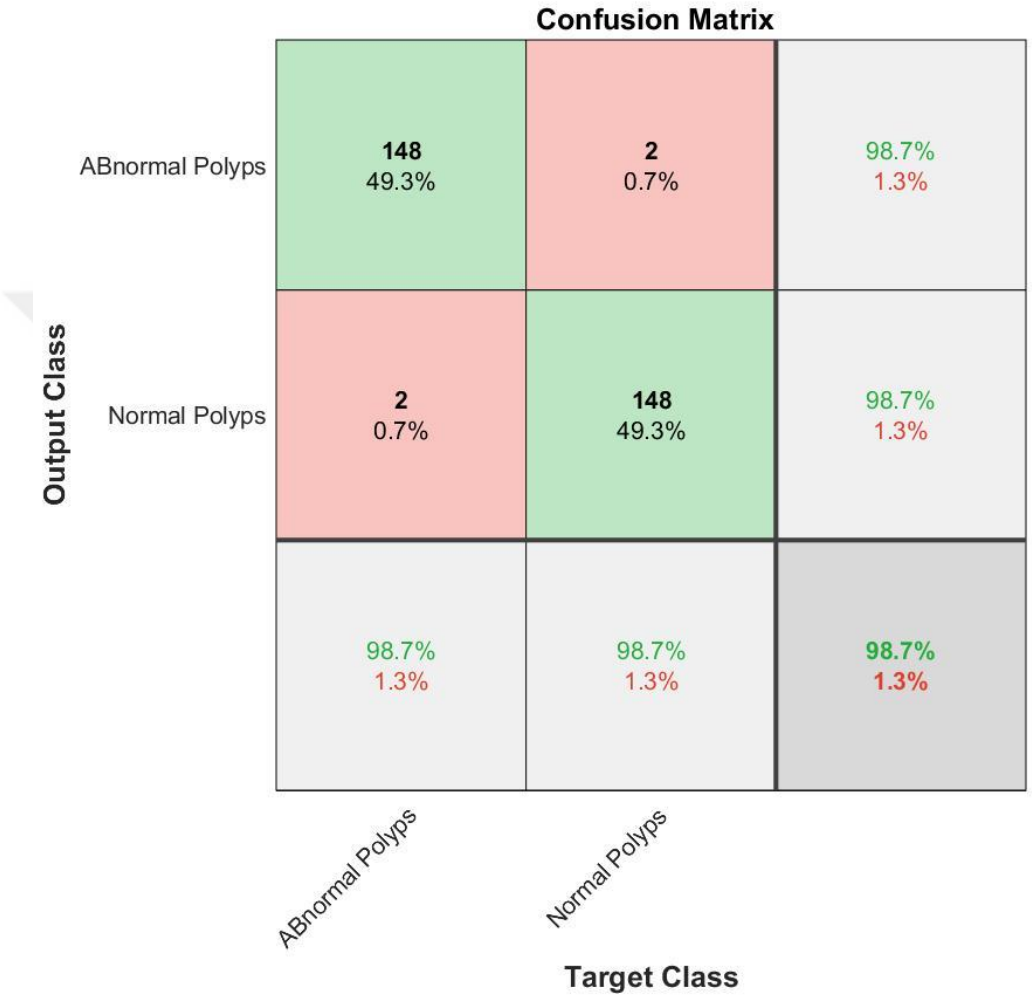

Figure 4.66 Our Method Results K1 Using SoftMax Classifier With Seven CNN Layers

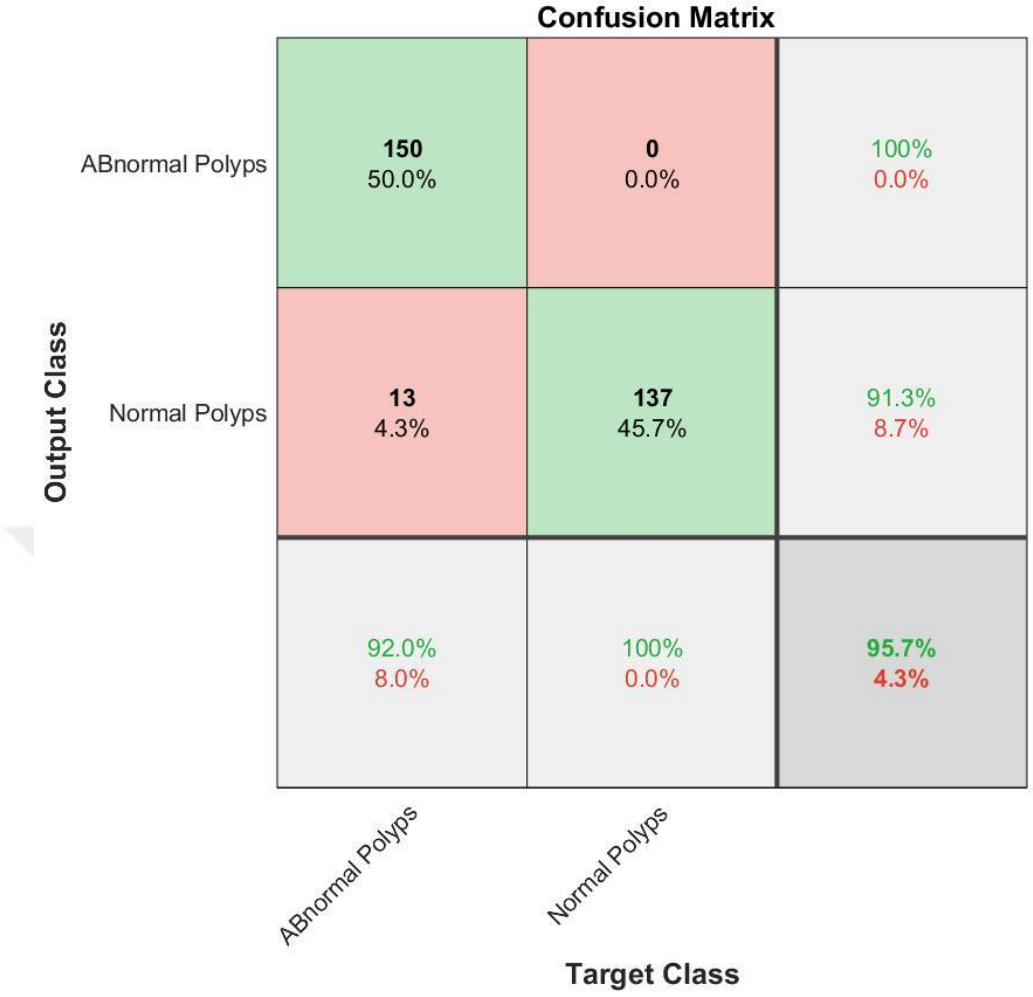

Figure 4.67 Our Method Results K2 Using SoftMax Classifier With Seven CNN Layers

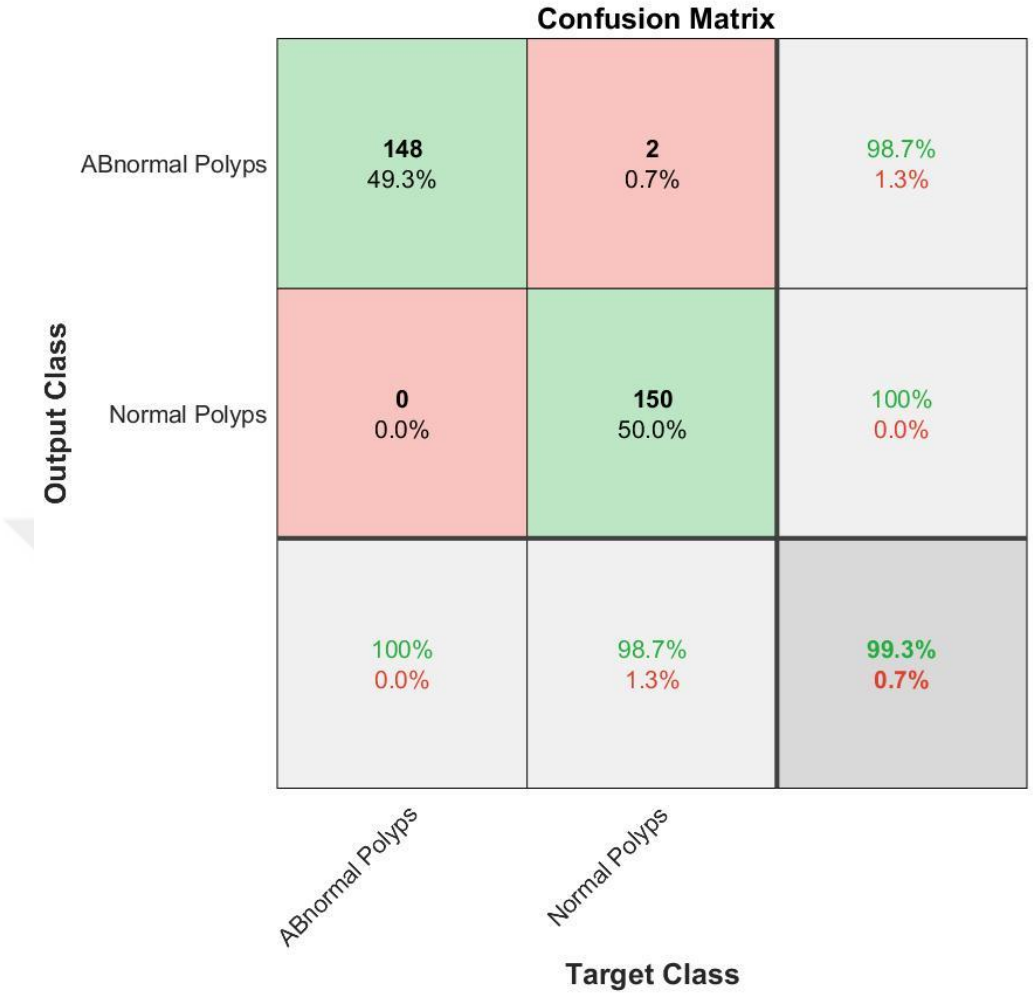

Figure 4.68 Our Method Results K3 Using SoftMax Classifier With Seven CNN Layers

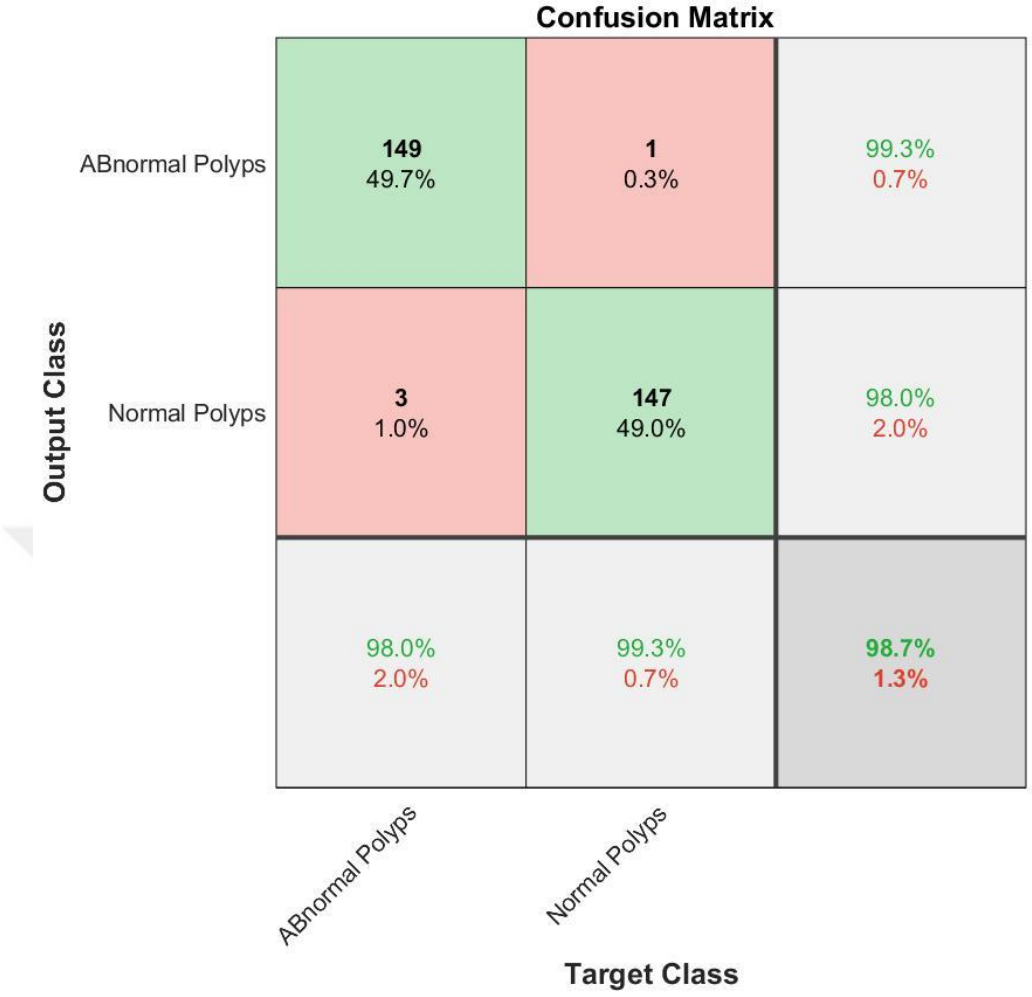

Figure 4.69 Our Method Results K4 Using SoftMax Classifier With Seven CNN Layers

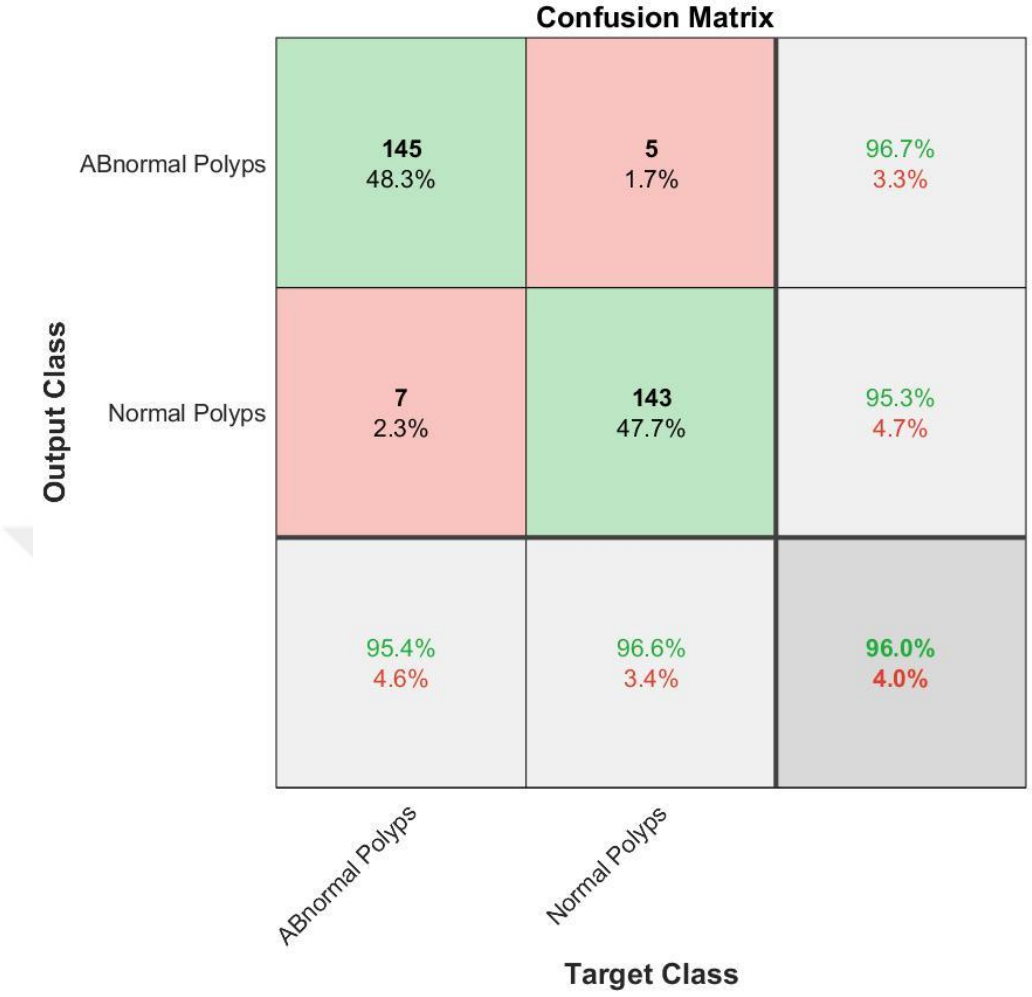

Figure 4.70 Our Method Results K5 Using SoftMax Classifier With Seven CNN Layers

Table 4.9 Seven CNN Layers Results

K-Fold	Test Results (%)	Time for Execution (Min)
K1	98.7	10:20
K2	95.7	9:12
K3	99.3	10:55
K4	98.7	10:48
K5	96.0	11:10
Average =	**97.68**	**10:29**

In the last stage, eight CNN layers are applied to the same problem, which is still not improving the result not like three CNN layers also the increasing of CNN layers mean increasing with time for execution and cost. Figure 4,71, 4.72, 4.73, 4.74, and figure 4.75 shows five K results and the table 4.10 shows the comparison between K and time for this model.

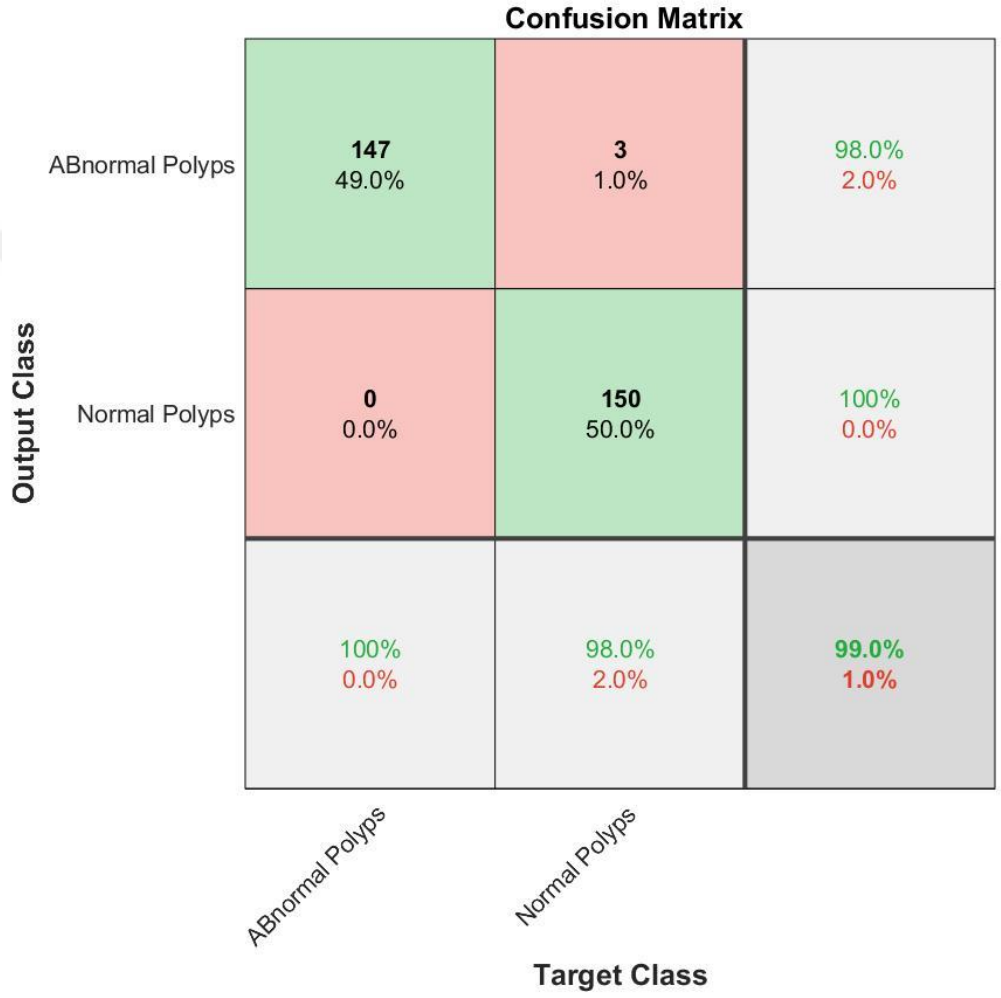

Figure 4.71 Our Method Results K1 Using SoftMax Classifier With Eight CNN Layers

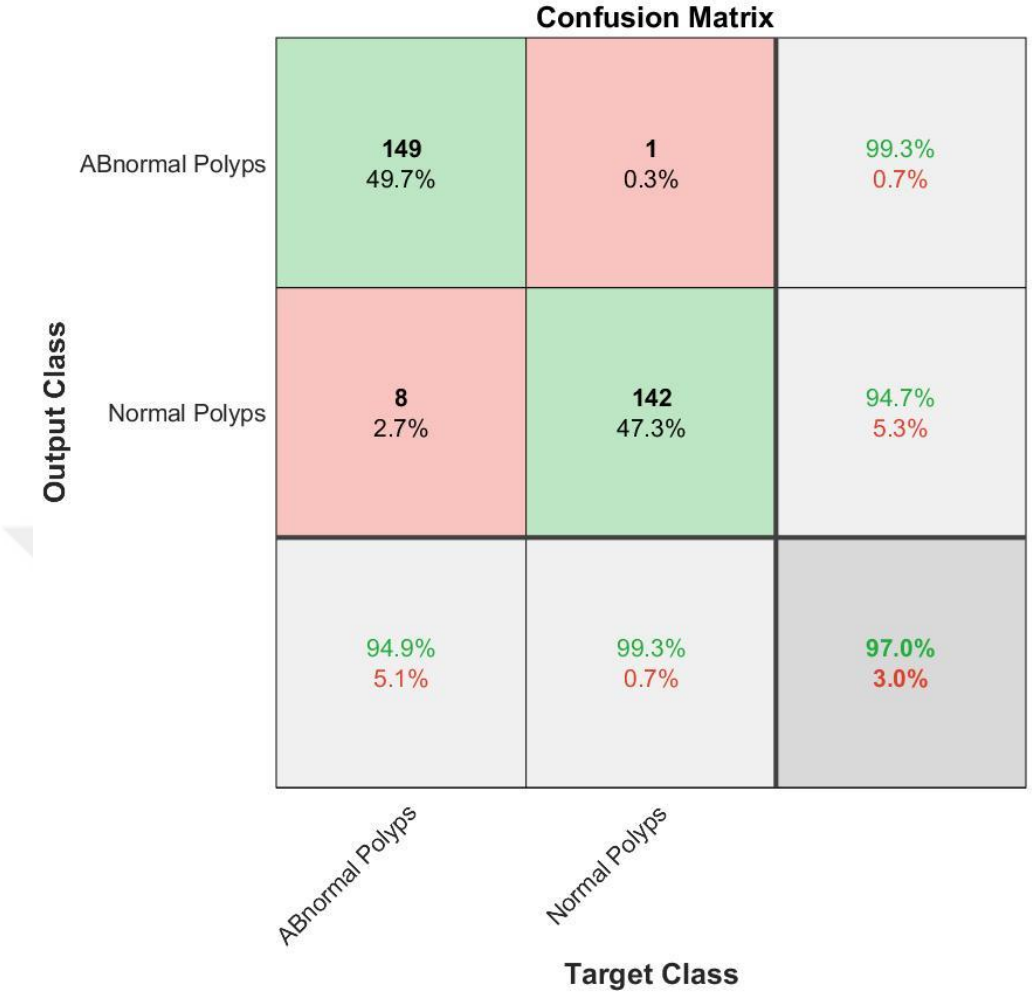

Figure 4.72 Our Method Results K2 Using SoftMax Classifier With Eight CNN Layers

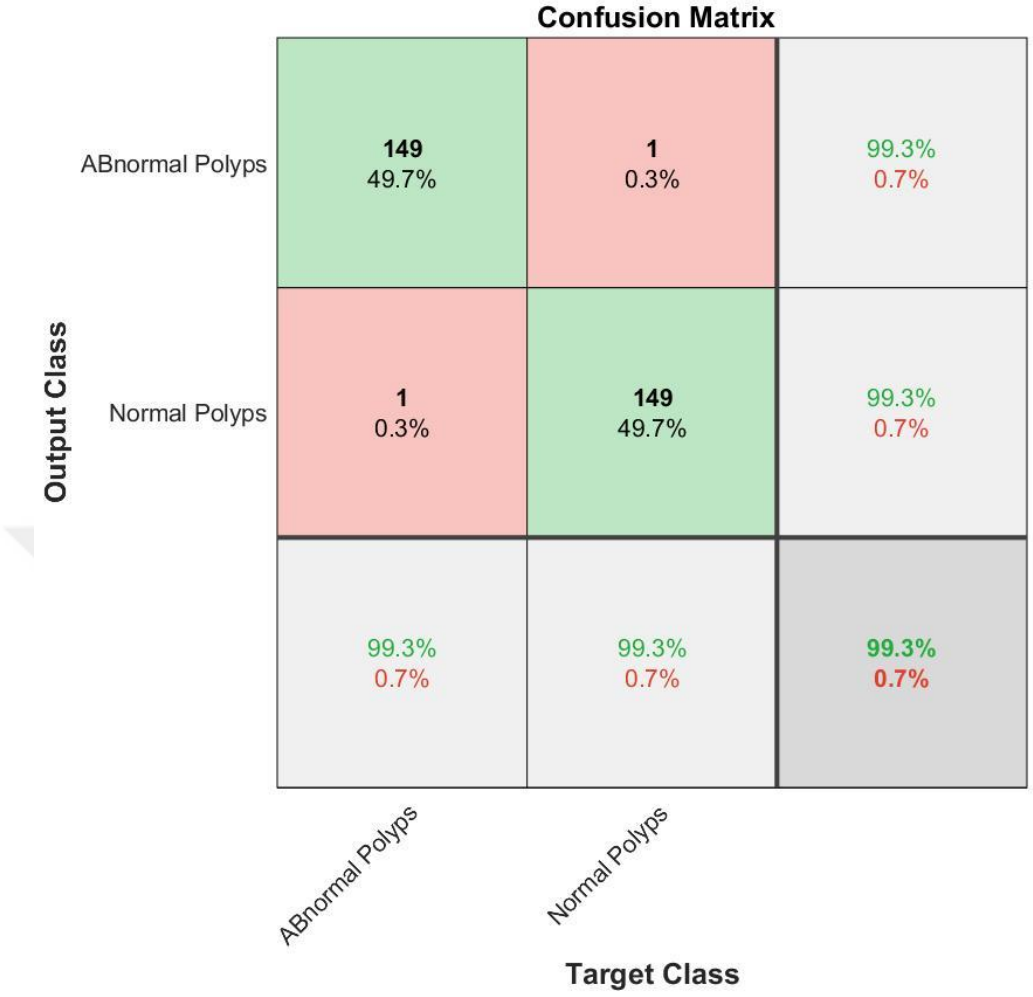

Figure 4.73 Our Method Results K3 Using SoftMax Classifier With Eight CNN Layers

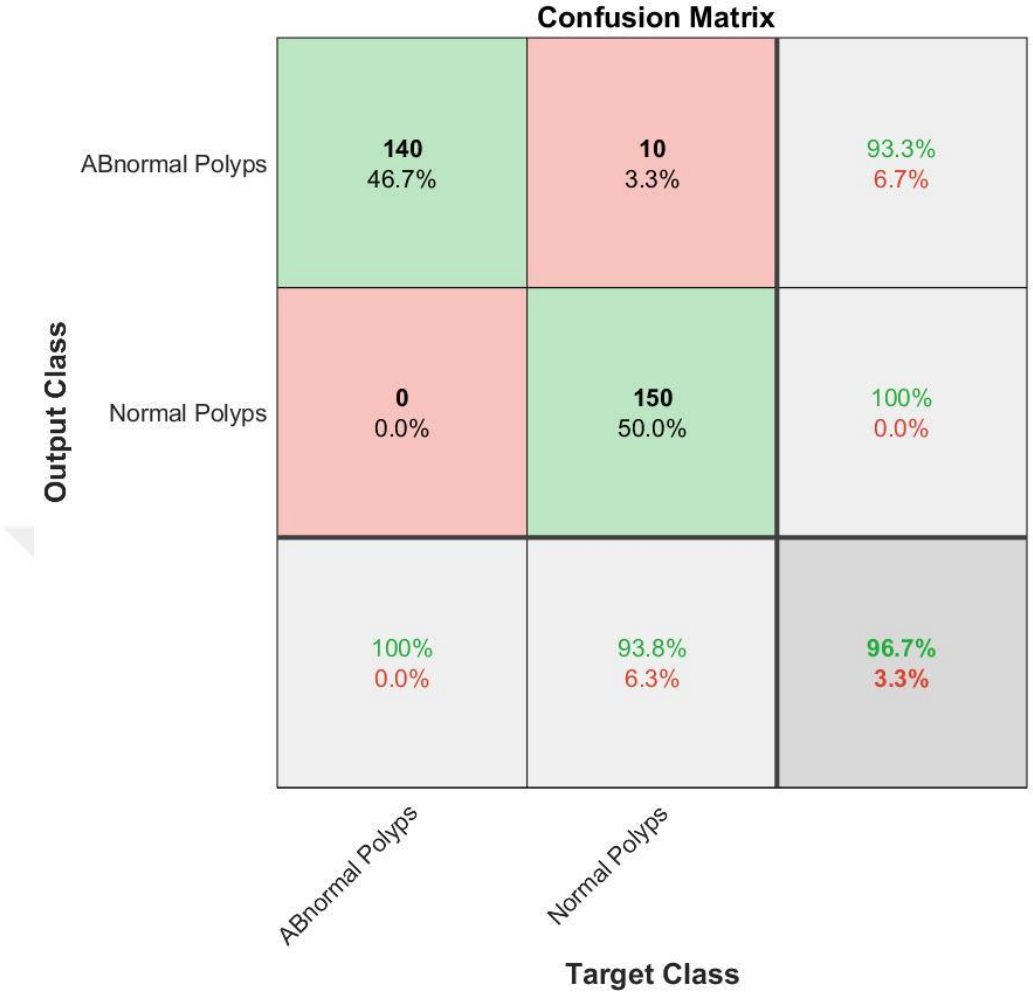

Figure 4.74 Our Method Results K4 Using SoftMax Classifier With Eight CNN Layers

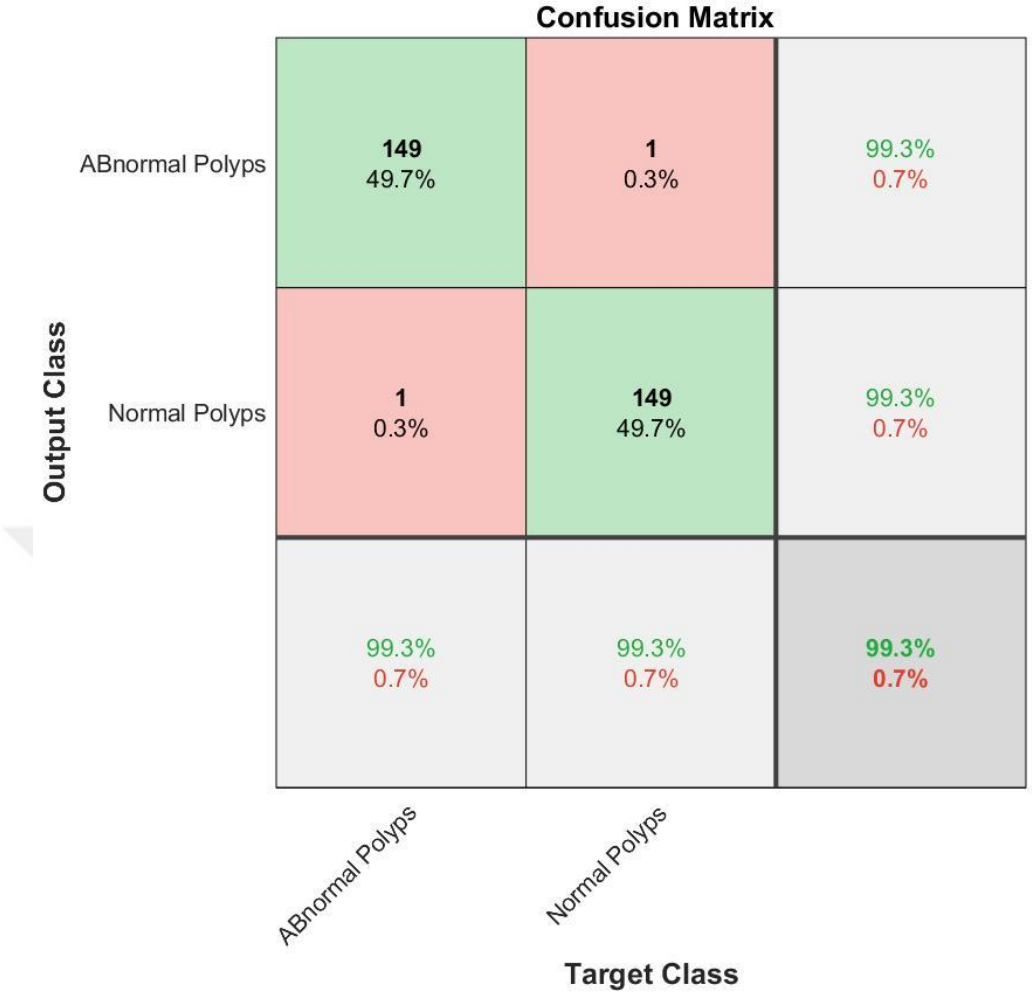

Figure 4.75 Our Method Results K5 Using SoftMax Classifier With Eight CNN Layers

Table 4.10 Eight CNN Layers Results

K-Fold	Test Results (%)	Time for Execution (Min)
K1	99.0	9:47
K2	97.0	9:34
K3	99.3	9:44
K4	96.7	9:37
K5	99.3	9:31
Average =	**98.26**	**9:39**

Instead of using SoftMax classifier in the three CNN layers, other type of classifiers are obtained for testing the results for classification. Two types of SVM applied in this study, Kernal and Linear. The Kernal function type with five fold cross validation applied and the confussion matrices of each K are presented in the figure 4.76, 4.77, 4.78, 4.79, and figure 4.80.

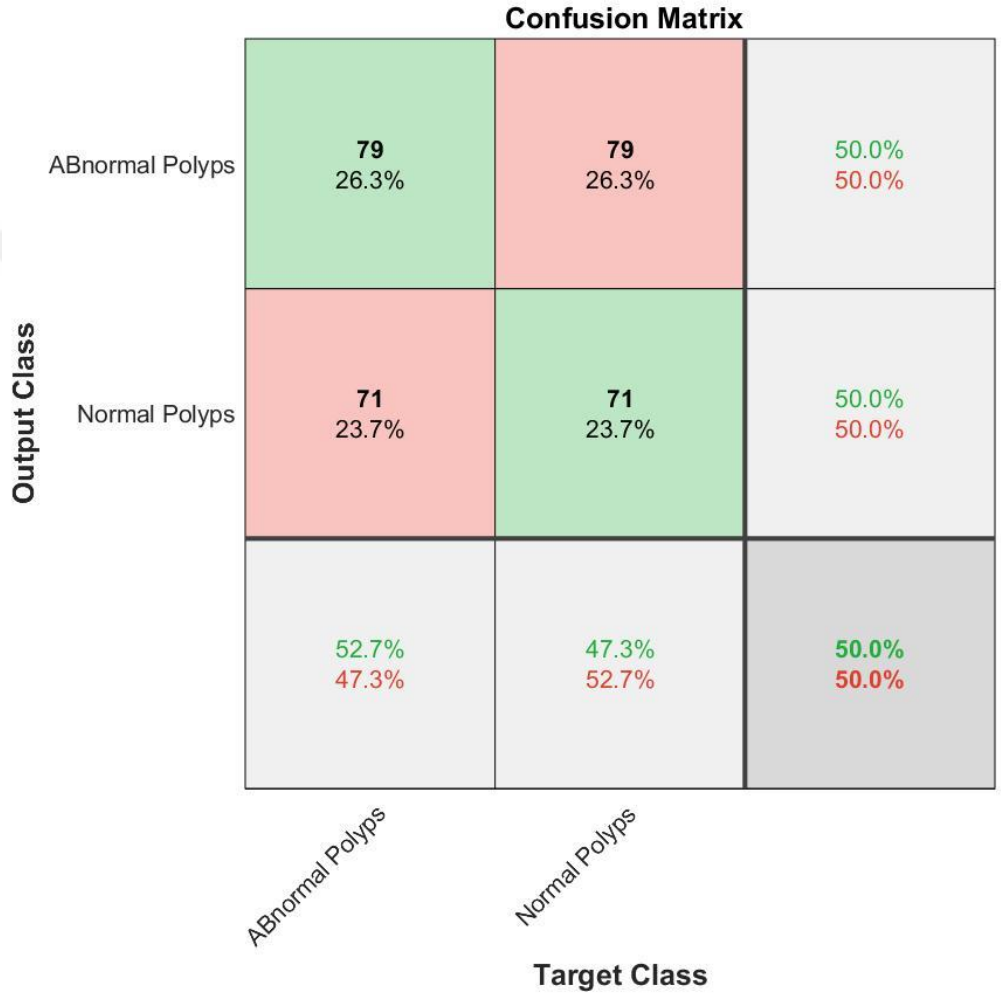

Figure 4.76 Our Method Results K1 Using Kernal SVM With Three CNN Layers

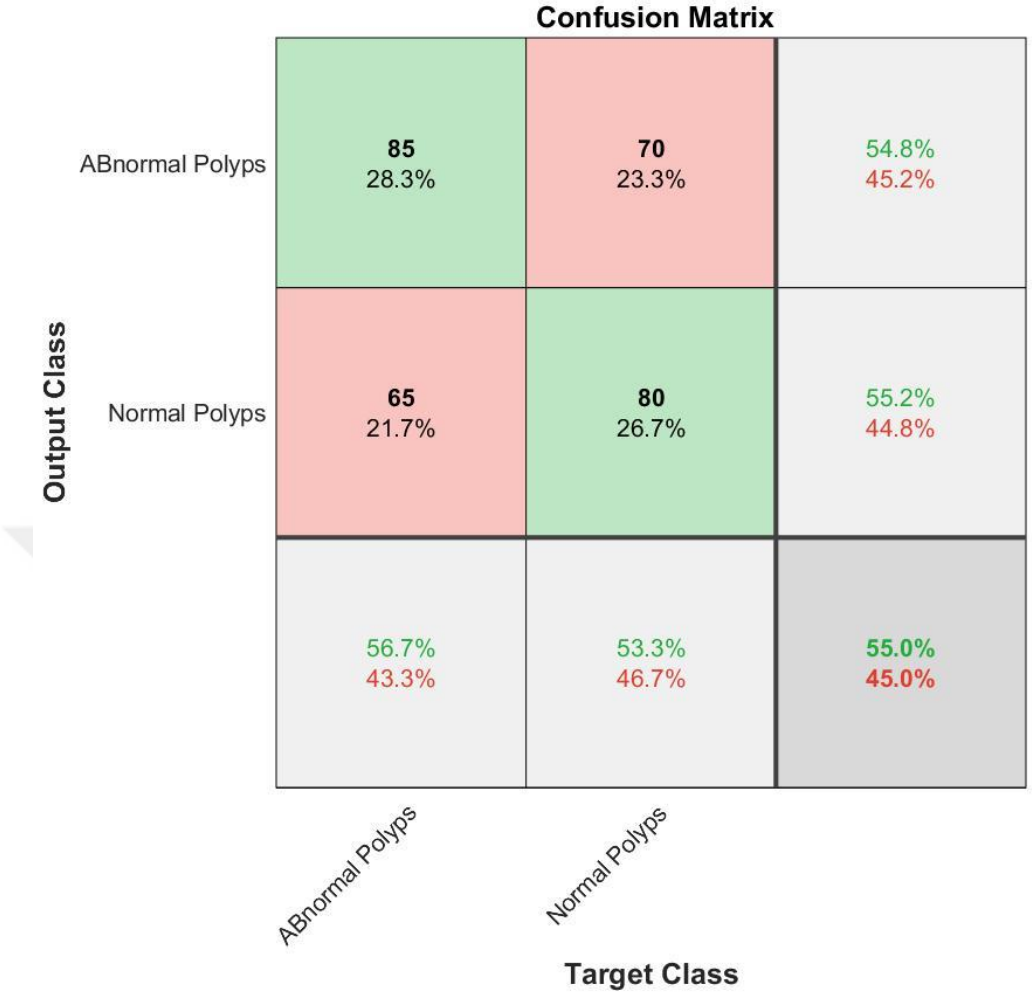

Figure 4.77 Our Method Results K2 Using Kernal SVM With Three CNN Layers

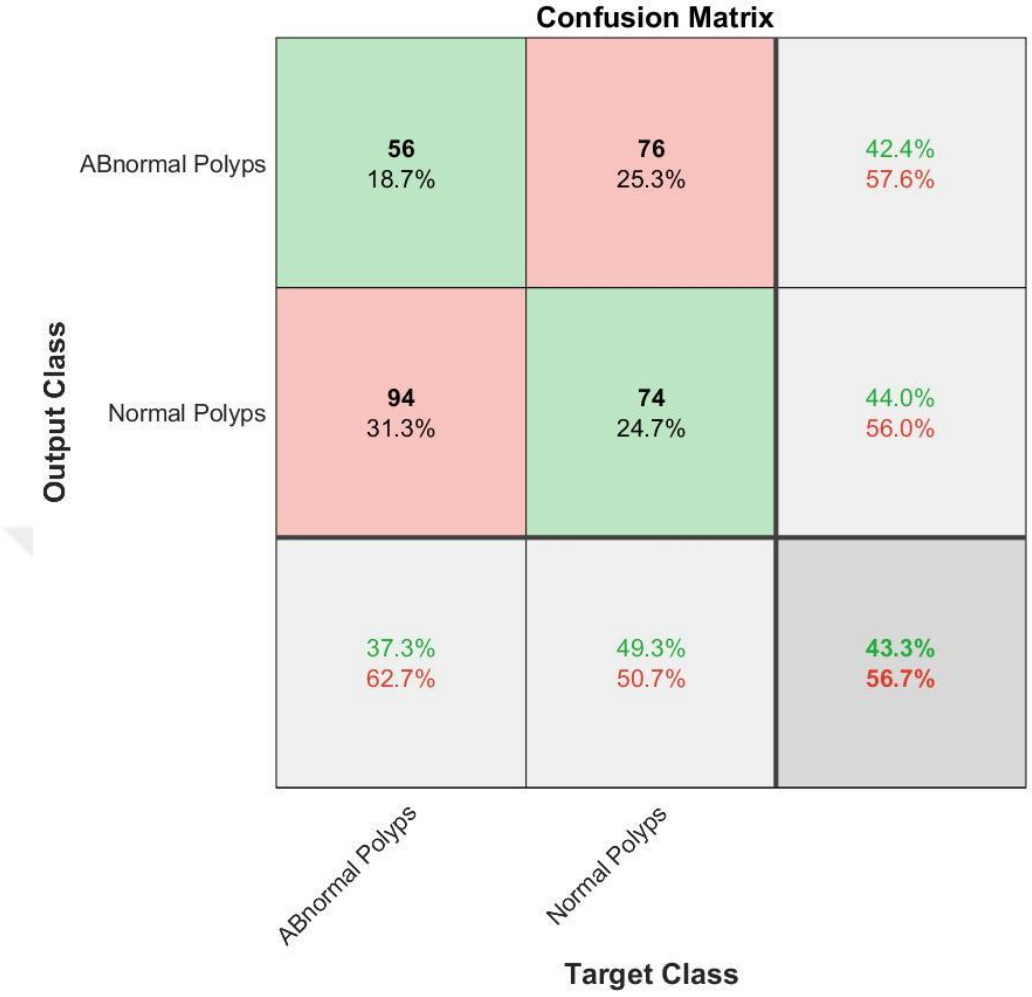

Figure 4.78 Our Method Results K3 Using Kernal SVM With Three CNN Layers

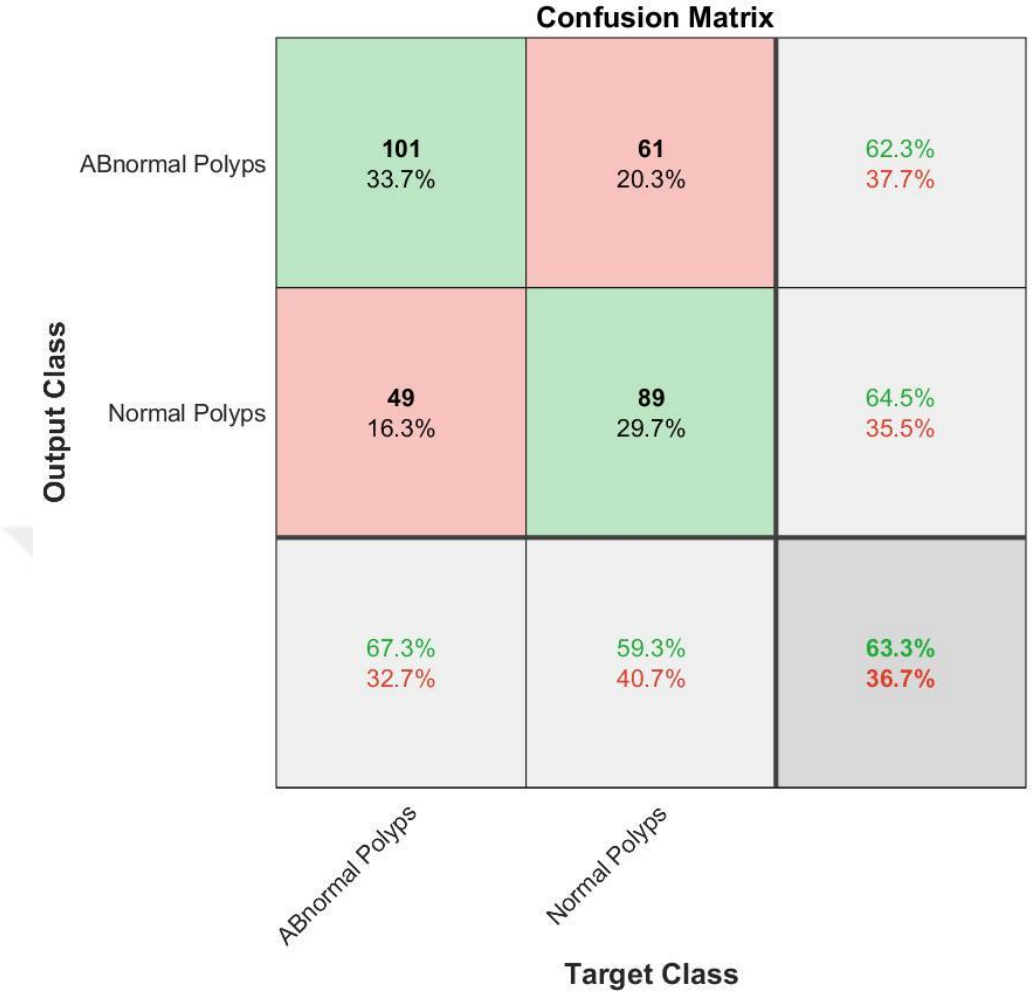

Figure 4.79 Our Method Results K4 Using Kernal SVM With Three CNN Layers

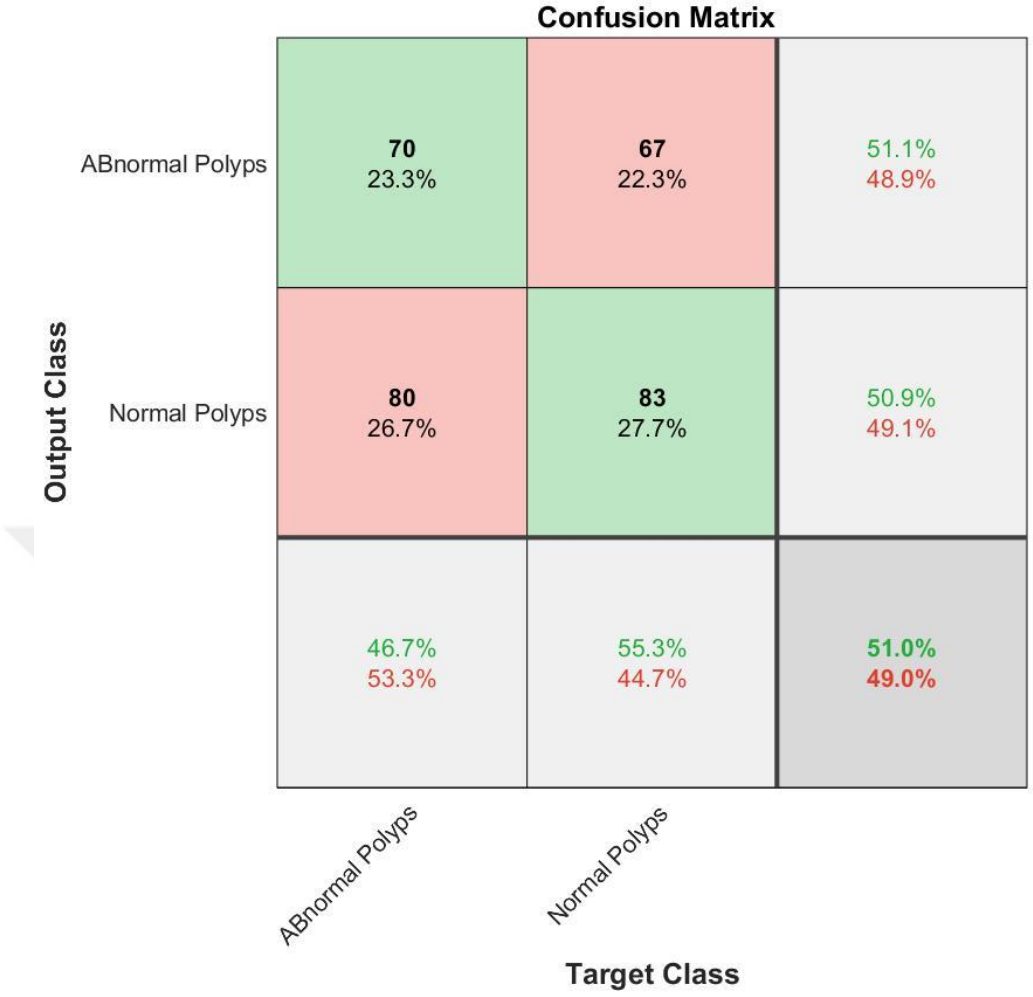

Figure 4.80 Our Method Results K5 Using Kernal SVM With Three CNN Layers

The Linear function type with five fold cross validation applied and the confussion matrices of each K are presented in the figure 4.81, 4.82, 4.83, 4.84, and figure 4.85. Table 4.11 shows the comparison of accuracy results between SoftMax classifier and two types of SVM classifiers (Kernel and Linear functions) with five of K-fold cross validation and three CNN layers.

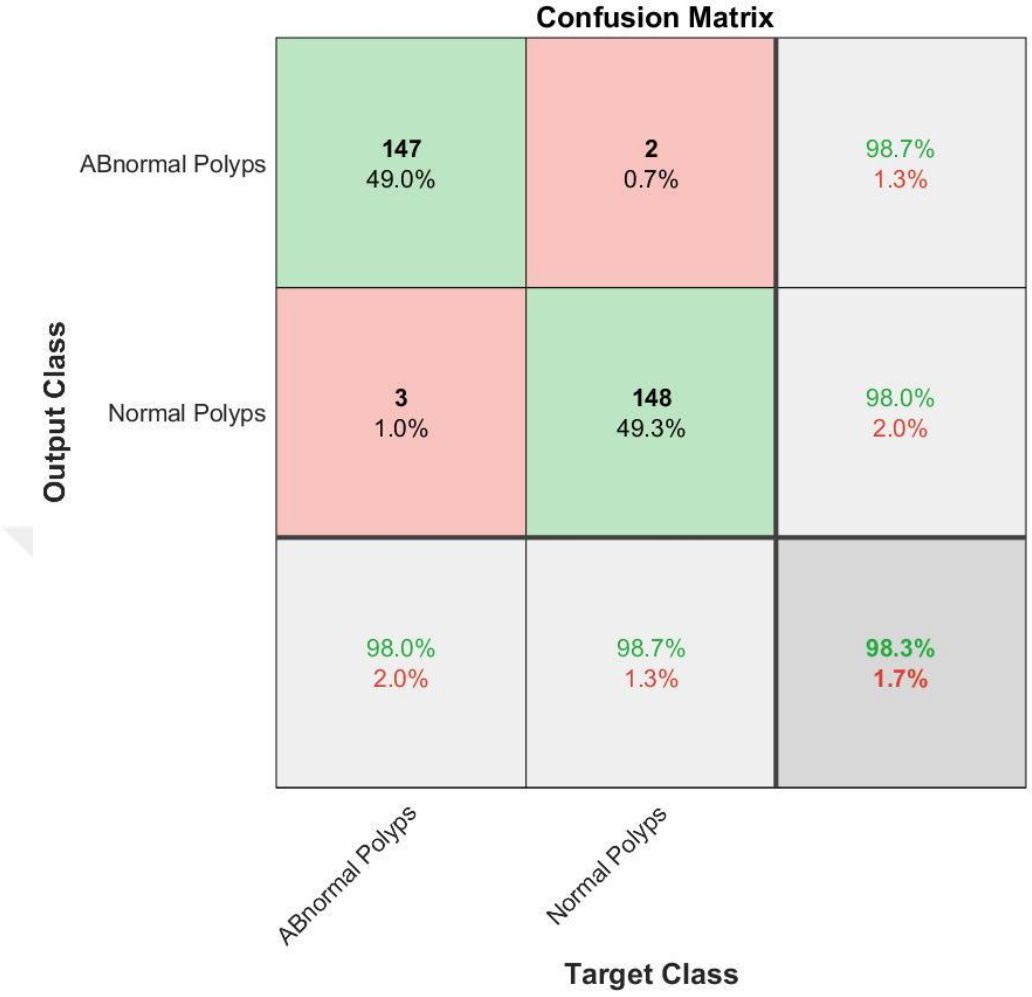

Figure 4.81 Our Method Results K1 Using Linear SVM With Three CNN Layers

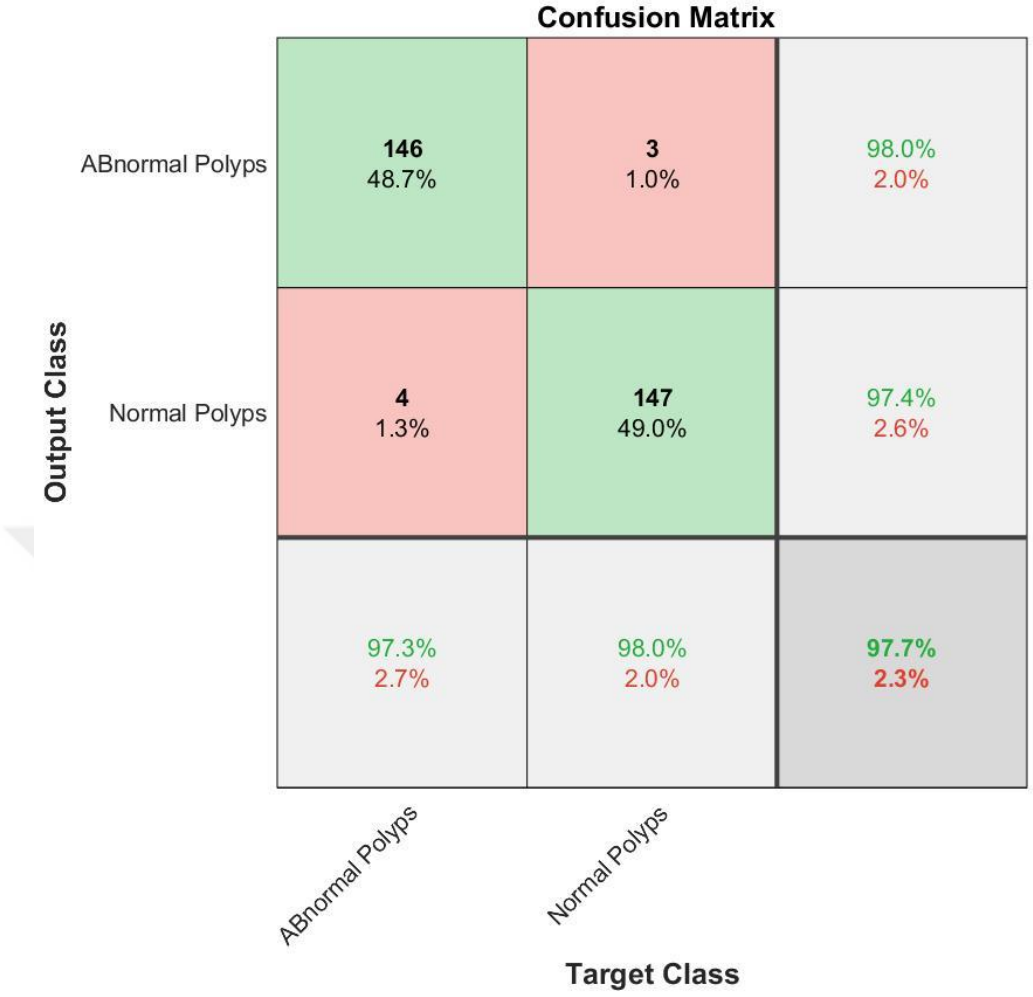

Figure 4.82 Our Method Results K2 Using Linear SVM With Three CNN Layers

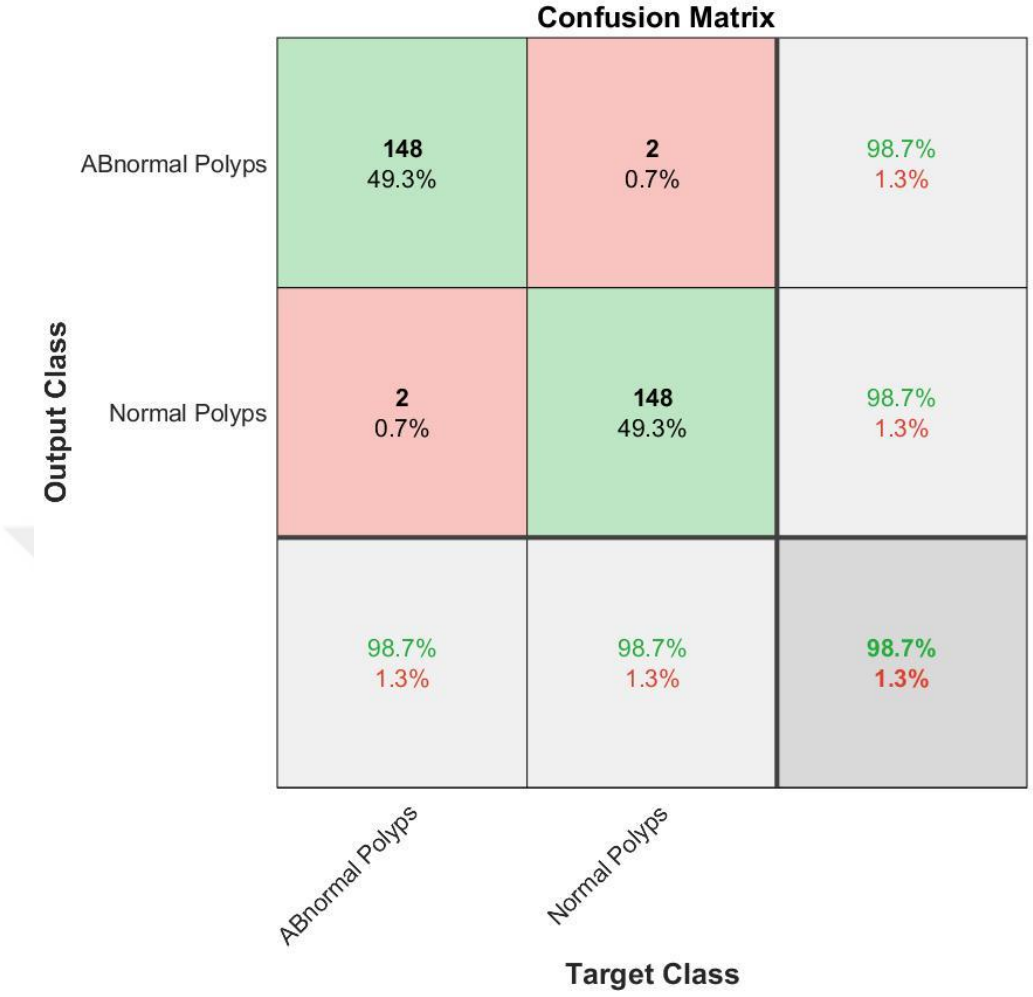

Figure 4.83 Our Method Results K3 Using Linear SVM With Three CNN Layers

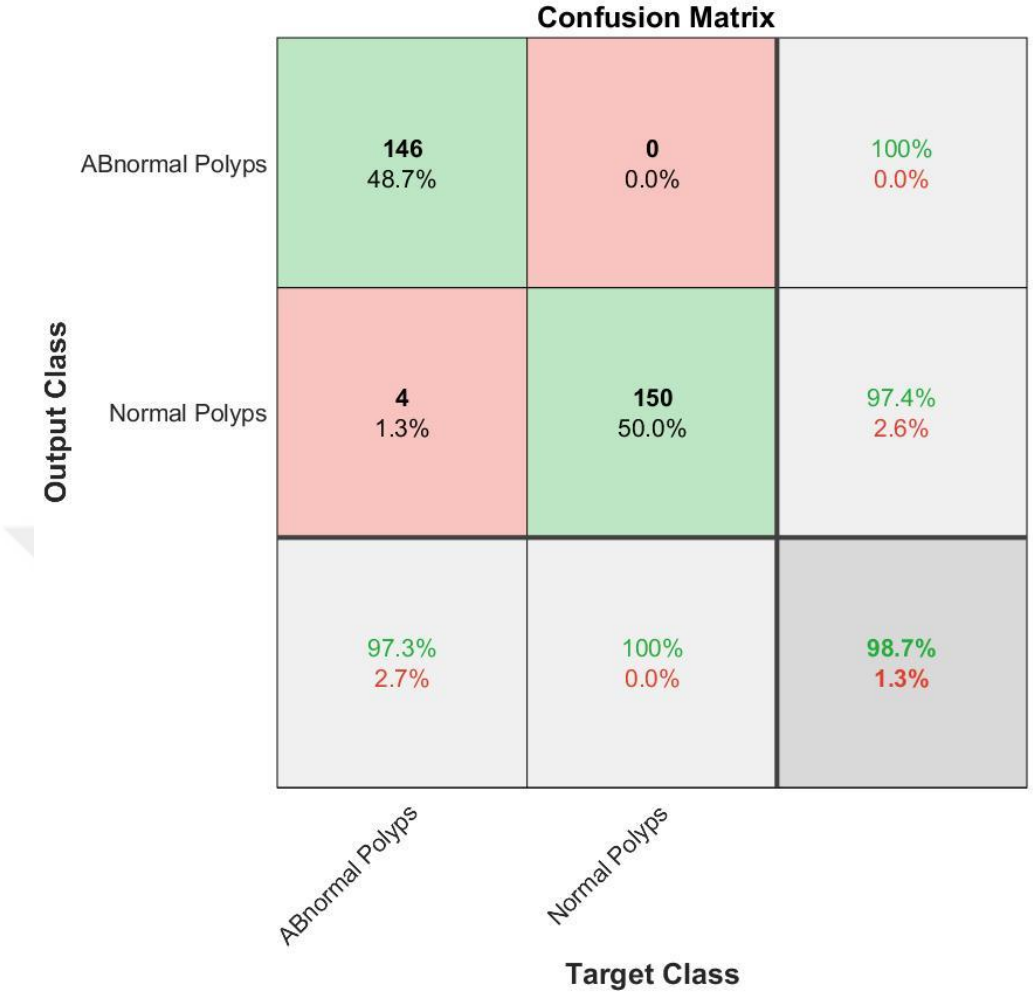

Figure 4.84 Our Method Results K4 Using Linear SVM With Three CNN Layers

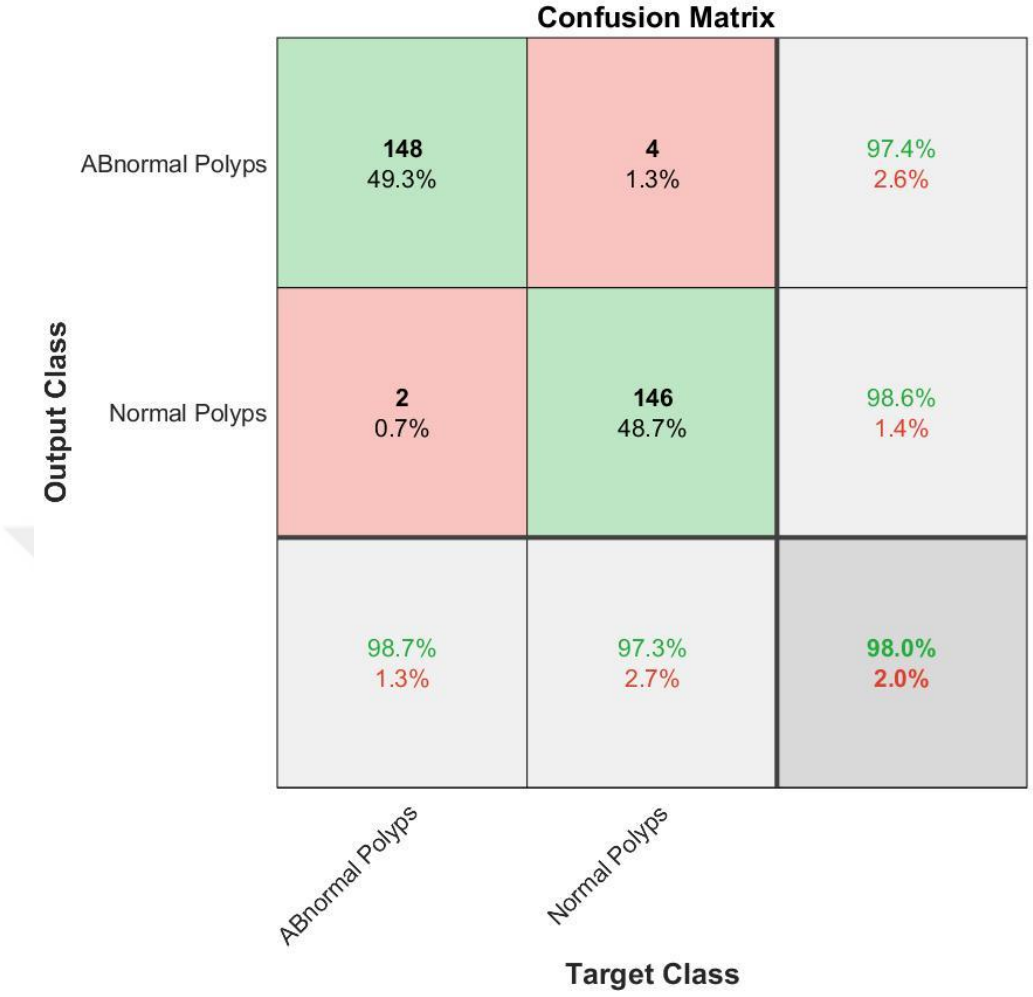

Figure 4.85 Our Method Results K5 Using Linear SVM With Three CNN Layers

Table 4.11: The Comparison Between Two Types of SVM and SoftMax Classifiers.

K-Fold	kernel Acc. (%)	Linear Acc. (%)	SoftMax Acc. (%)
K1	50.0	98.3	100
K2	55.0	97.7	99.7
K3	43.3	98.7	100
K4	63.3	98.7	100
K5	51.0	98.0	99.7
Average =	**52.52**	**98.28**	**99.88**

CHAPTER 5

CONCLUSION

In this thesis, efficient method proposed to detect Gastric cancer by using CNN. The proposed method extracted high level features and reduce the dimension of features. DenseNet201 presented 99.6% accuracy which is remarkable results when compared with other pre-trained networks. Our proposed network presented 99.88% which is best than all pre-trained networks. Furthermore, our method presented 03:05 min. time for execution less than other methods except AlexNet that presented 02:32 min.

Furthermore, we conclude that the applied of three layers of CNN for extracting features has best results which presented 99.88% for classification with SoftMax classifier, as we notice that the increasing of CNN layers which lead to increase with the time and cost complications.

The major difference of our method from other pre-trained methods is it consists from a smaller number of layers and presented high accuracy and presented fast execution time in most cases.

Automatic detection systems presented high cost saving and time consuming in different fields of medical image recognition and classification problems. The automatic detection systems are increased in the last years for different healthcare fields such as tumor detection, heart disease classifications, eyes disease classification and etc.

As future work the Gastric cancer can by classified by using another deep learning technique such as LSTM, Deep belief neural network and recurrent neural network. The use of deep learning techniques presented best results than traditional methods that used in previous studies. Therefore, the author advises to use different techniques of deep learning to increase the detection performance.

REFERENCES

[1] S. G. Fard, M. Taheri, Long non-coding RNA signature in gastric cancer, Experimental and Molecular Pathology, Volume 113, 2020, 104365, ISSN 0014-4800, https://doi.org/10.1016/j.yexmp.2019.104365.

[2] R. Xiang, X. Han, K. Ding, Z. Wu, CMIP promotes Herceptin resistance of HER2 positive gastric cancer cells, Pathology - Research and Practice, 2019, 152776, ISSN 0344-0338, https://doi.org/10.1016/j.prp.2019.152776.

[3] J. Wei, Z. Bu, Sentinel lymph node detection for gastric cancer: Promise or pitfall? Surgical Oncology, Volume 33, 2020, Pages 1-6, ISSN 0960-7404, https://doi.org/10.1016/j.suronc.2019.12.005.

[4] T. Chen, L. Zhao, S. Chen, B. Zheng, H. Chen, T. Zeng, H. Sun, S. Zhong, W. Wu, X. Lin, L. Wang, The curcumin analogue WZ35 affects glycolysis inhibition of gastric cancer cells through ROS-YAP-JNK pathway, Food and Chemical Toxicology, 2020, 111131, ISSN 0278-6915, https://doi.org/10.1016/j.fct.2020.111131.

[5] A. Kumar, R. Maskara, S. Maskara, I. J. Chiang, Conceptualization and application of an approach for designing healthcare software interfaces, Journal of Biomedical Informatics, Volume 49, 2014, Pages 171-186, ISSN 1532-0464, https://doi.org/10.1016/j.jbi.2014.02.007.

[6] G. Piho, J. Tepandi, D. Thompson, A. Woerner, M. Parman, Business Archetypes and Archetype Patterns from the HL7 RIM and open EHR RM Perspectives: Towards Interoperability and Evolution of Healthcare Models and Software Systems, Procedia Computer Science, Volume 63, 2015, Pages 553-560, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2015.08.384.

[7] Q. Ning, Z. Ma, X. Zhao, Deforms (KNN) - PseAAC: Detecting formylation sites from protein sequences using K-nearest neighbour algorithm via Chou's 5-step rule and pseudo components, Journal of Theoretical Biology, Volume 470, 2019, Pages 43-49, ISSN 0022-5193, https://doi.org/10.1016/j.jtbi.2019.03.011.

[8] Y. Tang, B. EM: A faster Bayesian linear regression algorithm without matrix inversions, Neurocomputing, Volume 378, 2020, Pages 435-440, ISSN 0925-2312, https://doi.org/10.1016/j.neucom.2019.10.061.

[9] C. N. Li, P. W. Ren, Y. H. Shao, Y. F. Ye, Y. R. Guo, Generalized elastic net Lp-norm nonparallel support vector machine, Engineering Applications of Artificial Intelligence, Volume 88, 2020, 103397, ISSN 0952-1976, https://doi.org/10.1016/j.engappai.2019.103397.

[10] S. B. Yang, T. L. Chen, Uncertain decision tree for bank marketing classification, Journal of Computational and Applied Mathematics, Volume 371, 2020, 112710, ISSN 0377-0427, https://doi.org/10.1016/j.cam.2020.112710.

[11] B. S. Olasege, S. Zhang, Q. Zhao, D. Liu, H. Sun, Q. Wang, P. Ma, Y. Pan, Genetic parameter estimates for body conformation traits using composite index, principal component, and factor analysis, Journal of Dairy Science, Volume 102, Issue 6, 2019, Pages 5219-5229, ISSN 0022-0302, https://doi.org/10.3168/jds.2018-15561.

[12] F. G. Tari, Z. Hashemi, Prioritized K-mean clustering hybrid GA for discounted fixed charge transportation problems, Computers & Industrial Engineering, Volume 126, 2018, Pages 63-74, ISSN 0360-8352, https://doi.org/10.1016/j.cie.2018.09.019.

[13] J. Ma, Y. Yuan, Dimension reduction of image deep feature using PCA, Journal of Visual Communication and Image Representation, Volume 63, 2019, 102578, ISSN 1047-3203, https://doi.org/10.1016/j.jvcir.2019.102578.

[14] T. Hirasawa, K. Aoyama, T. Tanimoto, S. Ishihara, S. Shichijo, T. Ozawa, T. Ohnishi, M. Fujishiro, K. Matsuo, J. Fujisaki, and T. Tada, "Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images," Gastric Cancer, vol. 21, no. 4, pp. 653–660, 2018.

[15] X. Zhang, F. Chen, T. Yu, J. An, Z. Huang, J. Liu, et al. (2019) Real-time gastric polyp detection using convolutional neural networks. PoLS ONE 14 (3): e0214133. https://doi.org/10.1371/journal.

[16] Y. Sakai, S. Takemoto, K. Hori, M. Nishimura, H. Ikematsu, T. Yano, & H. Yokota (2018). Automatic detection of early gastric cancer in endoscopic images using a transferring convolutional neural network. 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). doi: 10.1109/embc.2018.8513274.

[17] H. Wang, S. Ding, D. Wu, Y. Zhang & S. Yang (2018): Smart connected electronic gastroscope system for gastric cancer screening using multi-column convolutional neural networks, International Journal of Production Research, DOI: 10.1080/00207543.2018.1464232.

[18] B. Pushkar and M. Paul, "Early Diagnosis of Alzheimer' s Disease: A Multi- class Deep Learning Framework with Modified k- sparse Autoencoder Classification," pp. 1–6, 2016.

[19] J. Gu et al., "Recent advances in convolutional neural networks," Pattern Recognition., 2017.

[20] K. Ryczko, K. Mills, I. Luchak, C. Homenick, and I. Tamblyn, "Convolutional neural networks for atomistic systems," Comput. Mater. Sci., vol. 149, no. March, pp. 134–142, 2018.

[21] D. T. Tran, A. Iosifidis, and M. Gabbouj, "Improving efficiency in

convolutional neural networks with multilinear filters," Neural Networks, vol. 105, pp. 328–339, 2018.

[22] H. Wu and J. Zhao, “Deep convolutional neural network model based chemical process fault diagnosis,” Comput. Chem. Eng., vol. 115, pp. 185–197, 2018.

[23] H. Deng, L. Zhang, and X. Shu, “Feature memory-based deep recurrent neural network for language modelling,” Appl. Soft Comput. J., vol. 68, pp. 432–446, 2018.

[24] A. Gupta, A. T. Müller, B. J. H. Huisman, J. A. Fuchs, P. Schneider, and Schneider, G., “Generative Recurrent Networks for De Novo Drug Design,” Mol. Inform., vol. 37, no. 1, 2018.

[25] X. Song, Y. Liu, L. Xue, J. Wang, J. Zhang, J. Wang, L. Jiang, Z. Cheng, Time-series well performance prediction based on Long Short-Term Memory (LSTM) neural network model, Journal of Petroleum Science and Engineering, 2019, 106682, ISSN 0920-4105, https://doi.org/10.1016/j.petrol.2019.106682.

[26] C. Fan, Y. Sun, F. Xiao, J. Ma, D. Lee, J. Wang, Y. C. Tseng, Statistical investigations of transfer learning-based methodology for short-term building energy predictions, Applied Energy, Volume 262, 2020, 114499, ISSN 0306-2619, https://doi.org/10.1016/j.apenergy.2020.114499.

[27] Z. Kang, B. Yang, S. Yang, X. Fang, C. Zhao, Online transfer learning with multiple source domains for multi-class classification, Knowledge-Based Systems, 2019, 105149, ISSN 0950-7051, https://doi.org/10.1016/j.knosys.2019.105149.

[28] R. Wang, J. Xu, T. X. Han, Object instance detection with pruned Alexnet and extended training data, Signal Processing: Image Communication, Volume 70, 2019, Pages 145-156, ISSN 0923-5965, https://doi.org/10.1016/j.image.2018.09.013.

[29] Jahandad, S. M. Sam, K. Kamardin, N. N. A. Sjarif, N. Mohamed, Offline Signature Verification using Deep Learning Convolutional Neural Network (CNN) Architectures GoogleNet Inception-v1 and Inception-v3, Procedia Computer Science, Volume 161, 2019, Pages 475-483, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2019.11.147.

[30] C. Yang, Z. Liu, K. Liu, J. Zhong, Z. Han, A Loose Default Diagnosis Method for Oblique Bracing Wire in High-Speed Railway, IFAC-PapersOnLine, Volume 52, Issue 24, 2019, Pages 18-23, ISSN 2405-8963, https://doi.org/10.1016/j.ifacol.2019.12.370.

[31] Q. Zhou, Z. Zhou, C. Chen, G. Fan, G. Chen, H. Heng, J. Ji, Y. Dai, Grading of hepatocellular carcinoma using 3D SE-DenseNet in dynamic enhanced MR images, Computers in Biology and Medicine, Volume 107, 2019, Pages 47-57, ISSN 0010-4825, https://doi.org/10.1016/j.compbiomed.2019.01.026.

[32] I. Ha, H. Kim, S. Park, H. Kim, Image retrieval using BIM and features from pretrained VGG network for indoor localization, Building and Environment, Volume

140, 2018, Pages 23-31, ISSN 0360-1323, https://doi.org/10.1016/j.buildenv.2018.05.026.

[33] S. Tourdot, A. Scardino, E. Saloustrou, D.A. Gross, S. Pascolo, P. Cordopatis, et al., A general strategy to enhance immunogenicity of low-affinity HLA-A2. 1-associatedpeptides: implication in the identification of cryptic tumor epitopes, Eur. J. Immunol. 30 (12) (2000) 3411–3421.

[34] M. H. Andersen, L. O. Pedersen, J. C. Becker, P. T. Straten, Identification of a cytotoxic lymphocyte response to the apoptosis inhibitor protein surviving in cancer patients, Canc. Res. 61 (3) (2001) 869–872.

[35] Y. Tang, Z. Lin, B. Ni, J. Wei, J. Han, H. Wang, et al., An altered peptide ligand fornaive cytotoxic T lymphocyte epitope of TRP-2(180-188) enhanced im-munogenicity, Cancer Immunol. Immunother. 56 (3) (2007) 319–329.

[36] S. Zaremba, E. Barzaga, M. Zhu, N. Soares, K.Y. Tsang, J. Schlom, Identification of an enhancer agonist cytotoxic T lymphocyte peptide from human carcinoembryonic antigen, Canc. Res. 57 (20) (1997) 4570–4577.

[37]E. Garcia, R. Hermoza, C. B. Castanon, L. Cano, M. Castillo and C. Castanñeda, "Automatic Lymphocyte Detection on Gastric Cancer IHC Images Using Deep Learning," *2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS)*, Thessaloniki, 2017, pp. 200-204.

[38] G. Cao, W. Song and Z. Zhao, "Gastric Cancer Diagnosis with Mask R-CNN," *2019 11th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, Hangzhou, China, 2019, pp. 60-63.

[39] K. Pogorelov, K. R. Randel, C. Griwodz, S. L. Eskeland, T. de Lange, D. Johansen, C. Spampinato, D. T. D. Nguyen, M. Lux, P. T. Schmidt, M. Riegler, P. Halvorsen, Kvasir: A Multi-Class Image Dataset for Computer Aided Gastrointestinal Disease Detection, In MMSys'17 Proceedings of the 8th ACM on Multimedia Systems Conference (MMSYS), Pages 164-169 Taipei, Taiwan, June 20-23, 2017.