

אוניברסיטת בן-גוריון בנגב

Ben-Gurion University of the Negev

**הפקולטה למדעי ההנדסה**

**המחלקה להנדסת חשמל ומחשבים**

Faculty of Engineering Science

Dept. of Electrical and Computer Engineering

**פרויקט מסכם בקורס מבוא ללמידה עמוקה**

Introduction to Deep learning final project

חיזוי סרטן השד מסוג קרצינומה דוקטלית פולשנית לפי פריסת רקמות

# Predicting invasive ductal carcinoma in tissue slices

Group number: 50

Student #1: Name: Maxim Levit  
 ID: 310665419

Student #2: Name: Evgeny Andrachnik  
 ID: 317174084

Student #3: Name: Ziv Moldavsky  
 ID: 308992098

Project Instructor: Natalie Lang

Supervisor: Dr. Nir Shlezinger

T.A: Natalie Lang, Ben Wiesel

Submitting Date: 03/03/22

**Table of contents**

**1. Introduction 3**

**2. Goals 6**

**3. Dataset 10**

**4. Limitations 11**

**5. Infrastracture 16**

5.1. Folders hierarchy 23

**6. Data Analyze 16**

**7. Data pre-processing 16**

**8. Network and models 16**

**9. Performance and results 21**

**10. Benchmark 21**

**11. Result analysis 21**

**12. Major problems and solutions 23**

**13. Conclusions and Recommendations 28**

13.1. Project Summery 23

**14. References 30**

1. **Introduction**

Invasive Ductal Carcinoma (IDC) is the most common subtype of all breast cancers. Accurately identifying and categorizing breast cancer subtypes is an important clinical task, and automated methods can be used to save time and reduce error.

To assign an aggressiveness grade to a whole mount sample, pathologists typically focus on the regions which contain the IDC.

We are going to classify image patches and determine if it’s IDC or not IDC patch.

In order to do so, we based our project upon

1. **Goals**
2. **Dataset**

The original dataset consisted of 162 whole mount slide images of Breast Cancer (BCa) specimens scanned at 40x. From that, 277,524 patches of size 50 x 50 were extracted (198,738 IDC negative and 78,786 IDC positive). Each patch’s file name is of the format: uxXyYclassC.png — > example 10253idx5x1351y1101class0.png . Where u is the patient ID (10253idx5), X is the x-coordinate of where this patch was cropped from, Y is the y-coordinate of where this patch was cropped from, and C indicates the class where 0 is non-IDC and 1 is IDC. [3]

1. **Limitations**
2. **Infrastructure**

GPU: Quadro M4000

CPU:

* 1. Graphical user interface, text, application

     Description automatically generated**Folders hierarchy**

DeepLearningProjectNew – The main folder that contains all

Graphical user interface, text, application, email

Description automatically generatedTable

Description automatically generatedsdds

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text

Description automatically generatedTable

Description automatically generated

1. **Data Analyze**
2. **Data pre-processing**
3. **Network and Models**
4. **Performance and Results**
5. **Benchmark**
6. **Result analysis**
7. **Major problems and solutions (Difficulties and challenges)**

During our working progress we encountered some difficulties:

1. We had many image patches in our dataset.
2. **Conclusions and Recommendations**
3. **References**
4. *Our GitHub project.* By Ziv Moldavsky, Maxim Levit, Evgeny Andrachnik . Internet: <https://github.com/zivmo99/DeepLearningProjectNew> [Feb. 25, 2022].
5. *“Breast Histopathology PyTorch”* project. By Bucun. Internet: <https://www.kaggle.com/ambarish/breast-histopathology-pytorch> [Feb. 22, 2022].
6. *Data set: “Breast Histopathology Images”.* By Paul Mooney. Internet: <https://www.kaggle.com/paultimothymooney/breast-histopathology-images> [Feb. 22, 2022].
7. “|Breast Cancer” project. By Laura Fink. Internet: <https://www.kaggle.com/allunia/breast-cancer/notebook> [Feb. 22, 2022].
8. Internet: <https://gradient.run> [Feb. 22, 2022].
9. Cancer image classification based on DenseNet model. Ziliang Zhong et al 2020 J. Phys.: Conf. Ser. 1651 012143. Internet: <https://iopscience.iop.org/article/10.1088/1742-6596/1651/1/012143/pdf> [Feb. 22, 2022].