Table of Contents

[INTRODUCTION 2](#_Toc126313872)

[Week 1 report 3](#_Toc126313873)

[Week 2 report 4](#_Toc126313874)

[Week 3 report 5](#_Toc126313875)

[Week 4 report 6](#_Toc126313876)

[Week 5 report 7](#_Toc126313877)

[Conclusion 8](#_Toc126313878)

[References 9](#_Toc126313879)

# INTRODUCTION

The METLAIN LLP company is engaged in the development of technological machines, installations and the design of its own equipment or the assembly of existing installations that they order.

More recently, the company opened a data department and is actively developing such departments as Data Engineers and Data Scientist.

In order to gain experience for the development of a neural network, I was given free time to study articles provided by another organization.

# Week 1 report

Day 1

It all started with the collection of documents for the start of the internship, namely a trip to the clinic. Then a conversation with the staff and familiarization with the workplace and acquaintance with the project.

Day 2

I created requests to connect my PC to databases, to a Python program, then I got network access to the Internet.

Day 3

Started by contacting the medical center for a list of resources they could get to speed up the research work

Day 4-5

It was necessary to deal with sorting and more convenient arrangement of files and creating a table for more convenient study based on Python

# Week 2 report

Day 1-4

Cleaning data from unnecessary scientific papers that relate only to Data Science and Machine Learning

Day 5

Received feedback from the thesis supervisor to continue cleaning the list of scientific papers.

# Week 3 report

Day 1-3

As a result of data cleaning, we received about 100 research papers and, for more effective study, there was a distribution among the group members

Day 4-5

Based on the results, it was necessary to check the articles for availability in online resources to identify key segments for themselves

# Week 4 report

Day 1-3

We began to study the material in scientific articles 1-6, of which only one was significant in our project, the title of the article is Multi-Class Classification of Breast Cancer Using 6B-Net with Deep Feature Fusion and Selection Method, where this line played the main role: Proposed the method provides an average multi-class accuracy of 94.20% with a classification training time of 226 s in four breast cancer classes and an average multi-class accuracy of 90.10% with a classification training time of 147 s in eight classes. classes of breast cancer.

Day 4-5

Two scientific papers were studied namely: A scalable crowdsourcing approach and dataset for nucleus classification and segmentation in breast cancer; A Novel Multistage Transfer Learning for Ultrasound Breast Cancer Image Classification.

Briefly about each. According to the first scientific work, we can highlight: Using Mask R-CNN we will achieve high results, namely, we had a higher recall for sTIL (96.8 vs. 76.6). As for the second scientific work: According to the result, I will say that several models were carried out, and there is a table inside the article. Based on the same images as ours, we can achieve the same results.

# Week 5 report

Day 1-2

2 scientific papers were read namely: Construction and Validation of a Prognostic Risk Model for Triple-Negative Breast Cancer Based on Autophagy-Related Genes; Immune Effective Score as a Predictor of Response to Neoadjuvant Trastuzumab Therapy and a Prognostic Indicator for HER2-Positive Breast Cancer. About the first one, this is a classic risk model where certain parameters were summed up with the help of a neural network. A similar result can be said about the second one.

Day 3-5

At the end of reading all the scientific papers with the group, we collected his separate excel file where we wrote down the pros and the method that they used.

# Conclusion

As a result of the completion of the pre-diploma practice, in the luggage we have an accurate understanding of the principle of developing a neural network model for detecting breast cancer, using different algorithms, firstly, statistical analysis of variables, and secondly, which models are used in principle for this development segment. The resources that have been read have provided an understanding of the target variable and its role in our lives and provides a real world problem.

# References

Multi-Class Classification of Breast Cancer Using 6B-Net with Deep Feature Fusion and Selection Method: <https://www.mdpi.com/2075-4426/12/5/683/pdf>

A scalable crowdsourcing approach and dataset for nucleus classification and segmentation in breast cancer: <https://academic.oup.com/gigascience/article/doi/10.1093/gigascience/giac037/6586817?searchresult=1>

A Novel Multistage Transfer Learning for Ultrasound Breast Cancer Image Classification: <https://www.mdpi.com/2075-4418/12/1/135/pdf>

Construction and Validation of a Prognostic Risk Model for Triple-Negative Breast Cancer Based on Autophagy-Related Genes: <https://www.frontiersin.org/articles/10.3389/fonc.2022.829045/full>

Immune Effective Score as a Predictor of Response to Neoadjuvant Trastuzumab Therapy and a Prognostic Indicator for HER2-Positive Breast Cancer: https://www.mdpi.com/1718-7729/29/1/26/pdf