FOM - Hochschule für Oekonomie & Management

Hamburg

Master-Studiengang Big Data & Business Analytics

5. Semester

Development of a solution to detect skin lymphoma and to improve treatment processes

Betreuer: Prof. Dr. David Matusievicz

Autor: Jacqueline Franßen

Matrikel-Nr: 496804

5. Fachsemester

Hamburg, den 28.02.2021

Contents

List of Figures

List of Tables

1 Abstract

This scientific article focusses on the development of a neural network to detect tumors and to prevent skin cancer from recurring. The main purpose of this scientific article is to develop a model to be able to detect a specific type of skin cancer automatically. What is more, the mobile app should connect both patients and healthcare practitioners by digitalizing all measured data and saving them centrally on a secured database. The first business case is that patients are frequently reminded of going to a control appointment. Second, by using their smartphone's hardware, especially the camera can be used to take a photo of potential skin cancers.

2 Introduction

2.1 Problem statement

One of the most popular disease in our modern society is cancer. Many people die because of the consequences of tumors, evolving in their bodies and capturing nearly all of their good, healthy cells. The cause of tumors is the mutation of cells into 'bad' cells which clone themselves uncontrolled and irregularly. This makes it very hard to diagnose and prevent new tumors. Sometimes, even experts cannot diagnose precisely, e.g. in case of malignant tumors which are not delimited in the **mri!** (**mri!**) pictures but tend to proliferate in an unknown degree. Another problem is the point of time the diagnosis is made. Since many people who do not suffer from any chronic disease tend to go to the doctor infrequently, the tumor is be detected too late. What is more, most biopsies take up to multiple days which is a long time.

2.2 Aim and scope of this work

Given the above explained problems, the goals are to improve diagnostic and treatment of tumor patients by developing a model which predicts tumor metastasis and recurrence. Moreover, the system shall help doctors to quickly identify a patient whose cells are mutating so that he can intervene and treat the patient at the right time. Furthermore, the information process shall be improved by setting up a mobile and web application which keeps the patients up to date and serves as a communication base for healthcare providers.

The first aim of this scientific work is to develop a solution ... What is important, the developed model is only a reference model....

3 Fundamentals

3.1 Related Work

Deep learning methods use multiple layers of nonlinear processing units for feature extraction and transformation and to find deep relationships between complex variations under supervised and unsupervised procedures.

3.1.1 iarc! (iarc!)

The **iarc!** is a association with the aim to raise the development within cancer treatment and research. It was formed by the **who!** (**who!**) and has its main location in France. Besides, it provides a large library of cancerogens for users.
[1]

3.1.2 Ultrawideband, Stable Normal and Cancer Skin Tissue Phantoms for Millimeter-Wave Skin Cancer Imaging

[2]

3.1.3 Identification of 12 cancer types through genome deep learning

Sun et al. constructed ’14 models including 12 specific models, a total-specific model and a mixture model for cancer risk identification using a **dnn!** (**dnn!**) within a TensorFlow23 framework. We used an exponential decay method to optimize the learning

[1] **iarc˙world˙cancer˙report**
[2] **8344452**

rate, L2 regularization24 to minimize overfitting, and a sliding average model to increase the robustness of the model. For each specific model meant to identify a certain type of cancer, the detection accuracy, sensitivity and specificity are more than 97%, 98% and 97%, respectively. The mixture model, which is able to distinguish all 12 types of cancer, exhibited comparable performance. The total-specific and mixture models also demonstrated comparable performance. Using our model, cancerous tissue can be identified more conveniently and timely, thus providing an opportunity for earlier treatment. This approach to genome deep learning offers a new direction for disease diagnosis while providing a new method to predict traits based on genomic information.'[3]

Besides, Sun et al. mention multiple deep learning methods, such as 'Alpha Go' and object recognition which are used to recognize the locations of splice site promoters and enhancers in genetics. Furthermore, these are able to predict proteins secondary structure and functions.

Another advantage of using AI algorithms in cancer research is that they could achieve higher tumor-level sensitivity than pathologists. To give an example, today it is possible to automatically detect cerebral microbleeds from **mr!** (**mr!**) images which were identified by 3D **cnn!** (**cnn!**)s.

3.1.4 ham10000! (ham10000!)

Dermatoscopy TBD [4]

3.1.5 canscreen5! (canscreen5!)

The **iarc!** provides many projects based on different data sources. One large project is the **canscreen5!**[5] which collects all information on several cancer screening programmes across the globe. Thereby, the main goals are quality improvement and effective programme evaluation of cancer therapies. There are three types of cancer (breast, cervical and colorectal cancer) which are examined precisely. Besides, an overall graphic explains the distribution of cancer types across the worldwide popula-

[3] **sun˙identification**

[4] **ham10000**

[5] **iarc˙canscreen5˙project**

tion. Furthermore, the **canscreen5!** project provides analysis tools to be able to see the distribution of for example breast cancer in european countries.

3.2 Benchmarking

3.2.1 Yuma App

a mobile app to protect against sunburning

3.3 Different types of skin tumors

'The skin is made of a variety of cells, many of which are in constant motion. Round basal cells below the surface flatten as they rise to replace dead, flaking squamous cells on the surface. Melanocytes tan the skin in the sunlight, and Merkel cells give skin its ability to sense touch. When these cells become damaged, they may develop into skin cancer.' [6]

Tumor Name	Description	Benign/Malignant
Basal Cell Carcinoma	tbd	tbd
Recurrent Basal Cell Carcinoma	tbd	tbd
Squamous Cell Carcinoma	tbd	tbd
Melanoma	tbd	tbd
Karposi sarcoma (Rare Skin cancer)	tbd	tbd
Actinic keradosis (Rare Skin cancer)	tbd	tbd
Lymphoma of the skin (Rare Skin cancer)	tbd	tbd
Keratoacanthoma (Rare Skin cancer)	tbd	tbd
Merkel Cell Carcinoma	tbd	tbd

As described by **ndr!** (**ndr!**) [7], malignant skin lymphoma are caused by our immunesystem cells (also knows as lymphocytes). In such case, the lymphocytes are collected beneath the skin, grow uncontrolled and cause lymphomes. The reason for the mutation and why these cells cause skin lymphomes is unknown.

[6] **skin˙cancer˙types**
[7] **ndr˙lymphom**

3.3.1 Current diagnosis and therapy methods

When assuming skin lymphoma, doctors often examine the patient's blood or use X-ray and ultrasound to detect all tumor cells. After recognizing all malignant cells, there are several therapy methods, such as cortisone ointment, phototherapy, extra-corporeal photopheresis, X-rays, antibody therapy, chemotherapy and symptomatic therapy.

Given these numerous therapy methods, it is very important to diagnose the 'right' disease. Often, lymphoma tend to mutate and hide themselves as psoriasis and are treated like them for many years. But suddenly, the doctor notices that it is not a usual psoriasis and the lymphoma can immediately lead to the patient's death. For that reason, the diagnosis process should be repeated like a cycle and should include all different data sources (e.g. not only image recognition but also blood analysis) [8].

3.3.2 Stages of skin cancer

When diagnosing tumors, a often used scale is the **tnm!** (**tnm!**) grading which describes the current stage of the tumor. Tumor (t) describes the tumor's size, location and how deep it has grown into the skin. Node (n) indicates whether or not the cancer cells have spread to nearby lymph nodes or the channels connecting the lymph nodes. Metastasis (m) refers to whether the cancer cells have spread to distant organs.

Stage	Tumor devel
0	'Carcinoma in situ', tumor is only present in the epidermis (uppe
1	cancer is less than 2 centimeters, about 4-5 of an inch across, has not spread
2	cancer is larger than 2 centimeters across, and has not spread to nearby orga
3	cancer has spread into facial bones or 1 nea
4	cancer can be any size and has spread (metastasized) to 1 or more lymph nodes whic

According the the detected stage, the therapy is personalized. For instance, if tumors are in stage 1, a more simple treatment is administered whereas tumors in stage 4 have to be treated using more aggressive therapies, such as chemotherapies.

[8] **ndr˙lymphom**

Besides, there are additional criteria which can determine a higher risk for recurrence or spreading. These are for example a thickness greater than 2mm, invasion into the lower dermis or tiny nerves in the skin or a location on the ear or on a hair-bearing lip[9].

3.4 Current algorithms, solutions to recognize skin tumors and to predict metastasis

3.5 Overview: Apps to predict metastasis

There already exist some applications to predict certain types of cancer worldwide, such as [10]. These solutions provide several information about the tumors by showing them within many different types of charts. They do not specify on a certain patient (which would also cause problems with data privacy and protection) but explain them in general.
[11]

[9] **skin˙cancer˙types**
[10] **iarc˙predict˙cancer˙worldwide**
[11] **vijini˙gen˙alg**

4 Image Recognition and App Development

Here all used algorithms and patterns shall be explained

4.1 Image Recognition

4.2 Hybride App Development

5 Analysis and Development

5.1 Design Thinking Methods

In the following sections there are explained various kinds of design strategies to find out the real needs for the application being developed. These are adopted to two basic frameworks, the IBM Enterprise Design Thinking framework [1] as well as the Design Kit, proposed by IDEO.ORG [2]. Some of the strategies are modified a little bit and were implemented by a single person which can affect the objectivity and diversity of the methods. Nevertheless, they generated a highly usable product and many ideas for future development cycles.
In the given figure, a vertical letter shows the initially planned goals and features of the app. The X-Axis shows the time steps whereas the Y-Axis explains the complexity of tasks. The higher on task is mentioned, the more complex it is to realize.

5.1.1 Stakeholder Map

Figure **??** gives an overview of potential stakeholders. As can be seen in this figure, there are four groups of users: patients, doctors, the IT appartment of the hospital and the

[1] **ibm˙edt**
[2] **design˙kit**

5.1.2 Empathy Map

5.1.3 Hills

5.1.4 User Journey

5.2 User Research

5.2.1 User interviews

5.2.2 What is important? - Relevant features

5.2.3 How can processes be improved?

5.3 Development of system to predict metastasis and recidives

5.3.1 Challenges of development

Concerning the detection of skin moles are tattoos, skin lesions and tone of skin. In a first approach, these were not considered. But for future development, the usage of Lidar[3] sensor could be a solution. Since Lidar detects the 3D model of an object by calculating the distance from the detecting camera to the patient's skin, some special moles can be distinguished.

5.4 Data Research

5.4.1 Kaggle dataset

For training and developing the neural network, a basic Kaggle dataset was used [4]. It contains 3600 pictures of skin moles with the format 224x244 px. These 3600 files are divided into benign and malignant skin moles. For training, the neural network has to learn how to distinguish good from bad tumors and then set a warning. The

[3] **czichos˙introduction˙2018**
[4] **kaggle˙dataset**

dataset contains both training and test data. Since these are not labeled, in the first step, data preprocessing, labels have to be created.

5.5 crisp-dm! (crisp-dm!): Model Planning and Learning

5.6 U-net architecture to for medical imaging

A recent approach to detect tumors or objects that were marked by doctors in order to train a neural network is called 'U-Net Architecture' [5]. It is a **cnn!** architecture to 'fast and precise segment objects in images'. Figure XY gives an impression of a simple U-Net. 'Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations.'[6] One of the advantages is that during the first step (going right and downwards), objects are detected and in the second step (going right and upwards) the precise localization of the objects is calculated.
According to Ronneberger et al.[7], U-Net learns segmentation in an end-to-end setting. Furthermore, very few annotated images (approximately 30 per application) are needed. Touching objects of the same class.

5.7 Testing

5.8 Results

[5] **unet**
[6] **unet**
[7] **unet˙freiburg**

6 Results

6.1 Validation of results

6.2 Limitation in the development process

7 Conclusion and Outlook

7.1 Conclusion

7.2 Outlook

8 Abbreviations

CRISP-DM CRoss-Industry Standard Process for Data Mining

WHO World Health Organization

IARC International Agency for Research on Cancer

HAM10000 Human Against Machine with 10000 training images

MRI Magnetic Resonance Imaging

NDR Norddeutscher Rundfunk

TNM Tumor Node Metastasis

CanScreen5 Cancer Screening in Five Continents

CNN Convolutional Neural Network

DNN Deep Neural Network

MR Magnet Resonance

9 Appendix A

Ehrenwörtliche Erklärung

Hiermit versichere ich, dass die vorliegende Arbeit von mir selbstständig und ohne unerlaubte Hilfe angefertigt worden ist, insbesondere dass ich alle Stellen, die wörtlich oder annähernd wörtlich aus Veröffentlichungen entnommen sind, durch Zitate als solche gekennzeichnet habe. Ich versichere auch, dass die von mir eingereichte schriftliche Version mit der digitalen Version übereinstimmt. Weiterhin erkläre ich, dass die Arbeit in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde / Prüfungsstelle vorgelegen hat. Ich erkläre mich damit nicht einverstanden, dass die Arbeit der Öffentlichkeit zugänglich gemacht wird. Ich erkläre mich damit einverstanden, dass die Digitalversion dieser Arbeit zwecks Plagiatsprüfung auf die Server externer Anbieter hochgeladen werden darf. Die Plagiatsprüfung stellt keine Zurverfügungstellung für die Öffentlichkeit dar.

Ort, Datum (Vorname Nachname)

10 Appendix B