Universiteit Leiden
The Netherlands

Opleiding Informatica & Economie

Data-driven Prediction and Visualization of Rheumatoid Arthritis Flares

Heleen van Sonsbeek s1537806

Supervisors:

Dr. F. W. Takes & G.J. de Bruin

Advisors:

A. G. Pijl, S. Haitjema, O.A.G. Medina **(UMC Utrecht)**

BACHELOR THESIS

Leiden Institute of Advanced Computer Science (LIACS)
`www.liacs.leidenuniv.nl`

July 24, 2018

Abstract

Doctors determine the type of treatment for patients with rheumatoid arthritis based on the likelihood of whether a flare will occur. For doctors, this probability is very tough to estimate, which is why algorithms utilizing historic data have been developed to generate predictive models that help doctors decide on the most suitable treatment.

This research compares two techniques for generating such predictive models, namely machine learning and joint modelling. Although the joint modelling technique performs slightly better in terms of accuracy and performance, the machine learning techniques provide a much more insightful model which is easier to interpret. In addition to predictive models, a prototype intuitive, user-friendly visualization dashboard has been developed. It includes the results of these predictive models and in addition contributes to a better recording of necessary data, which will in turn help improve the developed models.

Acknowledgements

First of all I would like to thank my mother, who with her passion for this project has infected me with enthusiasm and which is why I wanted to be a part of this awesome, potential filled project. I want to thank Harry Pijl for granting me the opportunity to add a (small) contribution to this project and for guiding me throughout the entire process, as well as Saskia Haitjema. A lot of thanks to Marlies van der Goes for letting me experience the project from her side, for giving me a lot of useful insights and for all her help. I want to thank O'Jay Medina for being my mentor within the UMC Utrecht and for guiding and helping me with everything that I needed throughout this period.

Above all, I would like to give many thanks to my supervisors Frank Takes and Gerrit-Jan de Bruin, for their enduring support and faith in me, for helping me get the best out of myself and for providing me with excellent constructive feedback and knowledge that helped me to excel in this research.

Contents

Chapter 1

Introduction

Data analytics has been an explosively emerging field over the last years. It has created the possibility to generate useful insights from data that has been collected over the years to improve all kinds of organizations, and, with that, changing the way these organizations operate.

Data analytics is the science of examining raw data with the purpose of drawing conclusions about the information generated from this data [1]. Data is referred to as raw or unprocessed data, meaning unorganized facts such as characters, numbers and symbols. This is of little to no use for humans. Data needs to be processed, analyzed, organized or structured for it to be transformed into information, make it useful to humans [2]. Data analytics therefore gives great insights into the mountain of data collected by organizations.

Data analytics and big data analytics are terms that are often mixed in literature. However, these terms are not exactly the same. The data in data analytics is often referred to as structured data on a smaller scale, whereas big data is often referred to as unstructured data in high volumes. Therefore, since the data used for data analytics within the hospital is structured data, data analytics is the term that will be used for this research.

There are four types of data analytics, of which predictive analytics and descriptive analytics will be used in this research. Descriptive analytics focuses on the use of data to describe what has happened in the past, whereas predictive analytics uses data to predict what will happen in the future. Data-driven prediction therefore is predictive analytics done on structured data.

Data analytics offers a lot of advantages, like being able to personalize the customer experience and improving decision making, leading to overall cost reduction and a competitive advantage [3–6]. Because of these advantages, there has been a significant increase in the number of organizations making use of data analytics, big data analytics, or both, and the numbers keep getting bigger and bigger. In 2017, 53% of the organizations are making use of (big) data analytics, compared to only 17% in 2015 [7].

Even though the telecommunications sector and the financial services sector are fuelling the fastest adoption of (big) data analytics, this is a development that has not gone unnoticed in the health care sector either. As shown in Figure 1.1, around 57% of the organizations in the health care sector are making use of (big) data

analytics today, 38% of the organizations are aiming to do so in the future and merely 5% are not planning to do so [7].

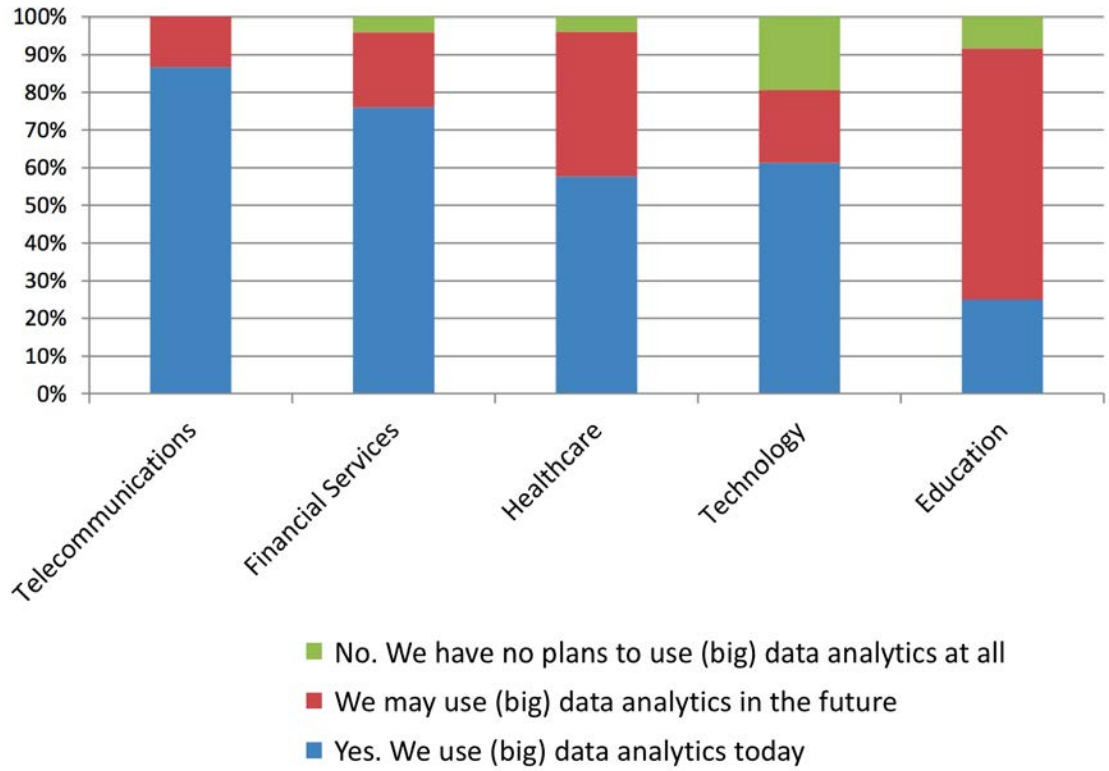

Figure 1.1: Adoption of (big) data analytics by vertical industry [7]

This is no coincidence. Many hospitals have recognized the advantages data analytics can give them, like being able to come up with daily and hourly predictions of how many patients are expected to be at each hospital, leading to reduced waiting times for patients and better quality of care, as done by four hospitals that make up the Assistance Publique-Hôpitaux de Paris [8].

Likewise, there has been a growing appetite in the Netherlands for making use of data analytics in Medicine in order to improve the health care. This brings us to the University Medical Center of Utrecht (UMC Utrecht), who have recently started a project called Applied Data Analytics in Medicine (ADAM) to implement data analytics with the goal of personalizing their health care system with a digitally supported hospital [9]. For this project, four departments of the hospital have been selected as pilot projects. These are neonatology, rheumatism, psychiatry and heart and vascular diseases [10].

The research conducted in this thesis surrounds the pilot project concerning rheumatism, specifically rheumatoid arthritis. Rheumatoid arthritis is an autoimmune disease, meaning that the body mistakes its own tissues as foreign and attacks them. The immune system attacks joints and other parts of the body, producing symptoms that often involve pain, fatigue and warm, swollen, inflamed-looking joints [30]. The symptoms of

this disease occur in flares, which mean that they occur all out of the sudden with fierce intensity.

The ultimate goal of the rheumatoid arthritis project surrounding this disease is to be able to predict the occurrence of such flares, with the use of different algorithms. To be optimally useful, these results must be properly visualized, meaning they must be integrated in a dashboard so that the rheumatism specialists can take these predictions into account while assessing patients.

The data available for this research is time-series and time-to-event data. Time-series, often referred to as longitudinal data, is data in which a response variable is measured at different time points, such as blood pressure, weight, or test scores measured over time. Time-to-event data, often referred to as survival data, on the other hand records times until an event of interest happens, such as times until a heart attack or in this case, times until a flare occurs [44].

This thesis contents the independent validation of already developed algorithms that do data-driven prediction by comparing it to standard machine learning algorithms, exploring the strengths and possible opportunities for improvement of the already developed algorithms that predicts rheumatoid arthritis flares. This is done so that the hospital will have an independent review of the algorithms developed, including strengths and weaknesses of the various techniques that can be used to decide which solution to use for extending to other similar diseases. In addition, this research will encompass the development of a prototype design dashboard for the rheumatism specialists, that has been created with the aim of improving efficiency in specialist consults, as well as contributing to better data collection. Furthermore, this study will focus on expansion possibilities that this and other research in this field can provide for similar diseases. The main research problem in this thesis therefore is *"Data-driven prediction and visualization of rheumatoid arthritis flares."*

To help generate insights into the results for this research investigation, three sub-questions have been defined:

1. How to develop a dashboard for rheumatism specialists that improves their efficiency, contributes to better data collection and encompasses the results of the models generated?
2. How do standard machine learning algorithms compare to the joint modelling approach?
3. How can this solution be extended to other similar problems within the hospital?

This thesis starts with an introduction to the project, the disease itself and a discussion of the various technologies used in this thesis in Chapter 2. After that, the dataset used for this research will be described in Chapter 3. Next, in Chapter 4, the exploratory dashboard visualization will be considered. Thereafter, the machine learning methods will be defined in Chapter 5. Consequently, the experiments, results and limitations of the research are discussed in Chapter 6 and finally a conclusion is drawn in Chapter 7, as well as the expansion possibilities.

Chapter 2

Background

This chapter will review the relevant literature and information related to the research conducted. Firstly, the background regarding the ADAM project will be discussed, in the second part an overview of rheumatoid arthritis as a disease will be given. Thirdly, an overview of joint modelling as a technique will be given, along with the implementation of this technique within the project and finally a description of the machine learning techniques and methods used in this research will be given.

2.1 ADAM

This chapter will review the relevant literature related to the project ADAM. First, a general overview of the project will be given, followed by an explanation of the background and aims for the project. Thereafter, the content of the four pilot projects will be briefly discussed, along with their progress. In addition, the vision for the future of the project will be considered.

The ADAM project comprises of four pilot projects, each of them being a data analytics project aimed at specific groups of patients within the knowledge domain of the focal points where health data can be deployed for giving personalized advises, personalized diagnostics, personalized treatment and personalized signalling [11].

2.1.1 Background and vision

ADAM was started in 2017 at the UMC Utrecht, with the goal of making better use of the data that the hospital has been collecting for years and years. The hospital has collected and stored all this data with the use of their Electronic Patient File, live-patient portal, the Research Data Platform and the eHealth pilots, and therefore, this data is available for analysis [10].

The decision to start this project came from the board of directors of the UMC Utrecht, who felt that, as seen in Figure 2.1, performing this project was in line with the vision that they hold as a hospital, namely

personalizing health care with a digitally supported hospital [9].

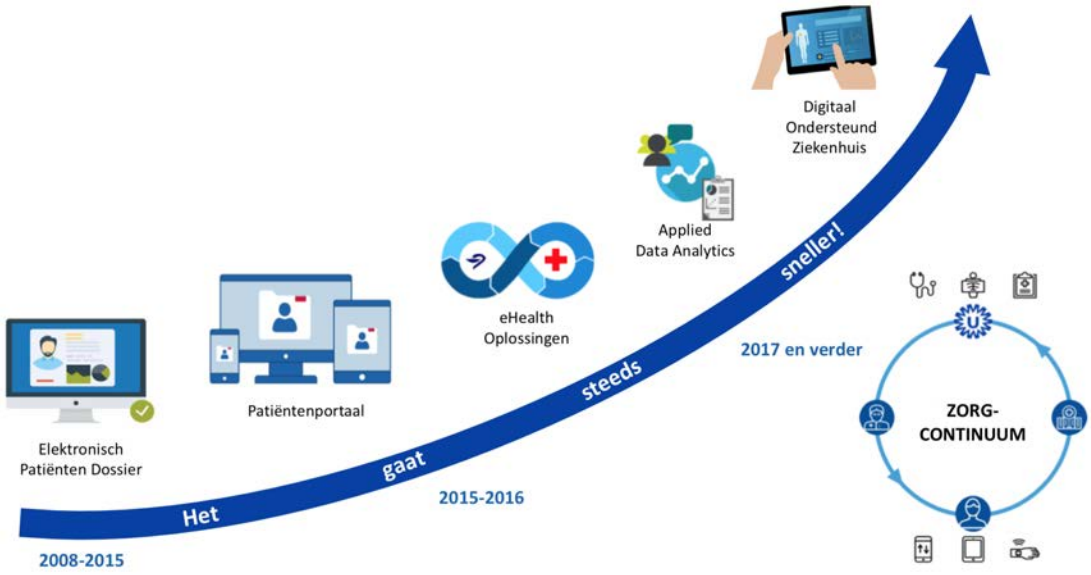

Figure 2.1: Vision held by the UMC Utrecht: personalizing health care with a digitally supported hospital. [9]

This project is seen as one of the key steps that contributes to this vision, since it allows for both implementing existing knowledge, with the use of simple algorithms, decision support and guideline adherence, and generating new knowledge, like pattern recognition, personalized prediction rules, automatic monitor analysis and business efficiency [10]. The insights created ultimately contribute to a higher quality health care system combined with a reduction in costs, as seen in Figure 2.2. It is stated that the ADAM project will contribute to improving customer satisfaction, generating quicker and more reliable diagnostics, therapy and monitoring and improving coordination within the care continuum, which will in turn improve the quality of the health care. Furthermore, the UMC Utrecht is convinced that the ADAM project will contribute to a reduction in waiting times and/or duration of stay for patients, an improved productivity of the clinicians and a reduction in costs for the treatment of chronical illnesses [9].

2.1.2 Scope and goals

The scope of the first phase of the ADAM project is as follows: four pilot projects should being selected, completed and implemented in practice. A division has been created in the hospital that serves as the core for this project and where in cooperation with external partners, the development of the data solutions take place.

For this phase of the project, the following goals have been established: all four pilot projects should be established and implemented in practice. More about these pilot projects can be found in Chapter 2.1.3. Another aim is to try out data-driven health care, with the focal point being that health care data can be implemented for generating personalized advises, personalized diagnosis, treatment and personalized

Figure 2.2: Results of the ADAM project as stated by the UMC Utrecht. [9]

signalling. Furthermore, a methodology should be established and tested for carrying out data-driven health care projects. In addition to that, the impact for the next phase concerning infrastructure, expected value for the patient and organization should be determined. Finally, thought formation should be started about the financing model and investments [10].

All areas of the hospital could participate in writing a plan of approach to be selected as one of the four pilot projects to be part of the ADAM project. The criteria of the pilot projects were as follows:

- Has to be directly applicable in the care for patients
- Has to be specific and has to have a short timeline
- Concern for the patient safety
- Availability of useable data
- Scaleable solutions aimed at a broader expansion within the health care
- Where possible coupled to the eHealth Care-continuum.

Based on these submitted project plans, the pilot projects were selected. The format of this pilot project can be found in Appendix E.

2.1.3 Pilot projects

After a lot of enthusiastic submissions, the following four pilot projects have been selected: neonatology, psychiatry, heart and vascular diseases and rheumatism, since the hospital felt that the most progress could be made in these areas. Below all four pilot projects are described briefly, along with their progress so far.

2.1.3.1 Neonatology

Neonatology is the department of pediatrics that involves the medical care of newborn infants, with a special focus on the ill or prematurely newborn. This specialty is based in the hospital, and is usually practiced in neonatal intensive care units (NICUs), which are special intensive care units designed to care for these ill or prematurely newborn infants.

Doctors practising this speciality want to know as quickly as possible whether and when which prematurely born babies get sick as a result of a sepsis, which is a type of blood poisoning. Furthermore, they want to prevent the administration of unnecessary antibiotics, and therefore want to know as soon as possible whether or not the administration of certain antibiotics might be necessary or not. The ultimate goal of the project is to know half a day earlier whether or not a sepsis is occurring.

This leads to the following questions to be answered in this pilot project, with the results to be as follows after one year:

1. **Is it possible to predict the probability of neonatal late-onset sepsis for prematurely born babies (<32 weeks).**
   A sepsis has been detected 12 hours earlier with a 70% certainty

2. **Is it possible to predict the type of bacteria (grams positive/negative) at the suspicion of late-onset sepsis**
   A positive bloodtest and the right gram coloring has been predicted with 90% certainty [12] [10].

2.1.3.2 Psychiatry

Psychiatry is the department that deals with the diagnosis, prevention and treatment of mental disorders. Since the psychiatry field is a very broad one, a choice has been made to focus on one particular element of this practice, namely psychosis. A psychosis is an abnormal condition of the mind that results in difficulties telling what is real and what is not. Symptoms of this illness include false beliefs and seeing or hearing things that others do not see or hear [13].

Doctors involved in this speciality want to know which medicine is immediately effective with patients battling this disease, since they want to prevent unnecessary medication with often unpleasant side effects. Doctors want to prescribe a medicine tailored to the specific patient, improving the loyalty in therapy, shorten the treatment time and eventually will lead to a better quality of life, both for the patient as well as their family.

A challenge within this project is that the effect of the medication has not been uniformly recorded in the data, which is why the model does not predict accurately enough. Therefore a dashboard has been created to enhance this uniform registration of the data, so that the ultimate goal of the project, namely for psychiatrists to give better advice about the use of anti-psychotics, can be achieved [10] [12].

This leads to the following question to be answered in this pilot project, along with the result that has been achieved after the past year:

Is it possible to predict which anti-psychotic medicine will work (or not) for an individual patient who needs to receive medication as well as to have an earlier insight into the effects and side-effects of the medication.
Insight in use of medication, graphical display through a dashboard [10].

Unfortunately, the data turned out to be unuseful at this moment, which led the relevant stakeholders to decide to develop a dashboard to achieve one of the goals, as well as to contribute to better aquiring of data.

2.1.3.3 Heart and vascular diseases

Heart and vascular diseases is the discipline within the hospital that concerns complications occurring to the heart and arteries of patients as a result of a substance called plaque that builds up in the walls of the arteries. This buildup narrows the arteries, making it harder for blood to flow through. If a blood clot forms, it can stop the blood flow. This can cause a heart attack or stroke [14].

A lot of data has been collected concerning this area all over the world, resulting in proof that certain factors will heavily increase the chance of the consequences above occurring. Unfortunately, most of these risks are not known to patients to such a degree, while, if they would be aware of this, actions can be taken that could heavily decrease the chances of such things happening.

This is what triggered the Heart and Vascular diseases specialists to come up with a solution for cardio-vascular risk management, which is to have a dashboard developed that shows all risk factors per patient, coupled to a personal risk score of a heart or vascular disease within the next ten years. Furthermore, it shows what the treatment and adjustment in lifestyle could give a patient for his or her health [12]. This contributes to the ultimate goal of the project, namely to have patients and doctors decide together which preventative therapy would be best, something that will increase convenience for the professional, patient empowerment and therapy loyalty [10].

This lead to the following question to be answered during the pilot project:

What is the optimal visualisation (dashboard) of risk profiles and scores for the individual patient (and how can this be integrated in the EPD)?

The results of this after one year are as follows:
Determining the personal cardiovascular profile with the use of an interactive dashboard in HiX, with effects of therapy on ten years risk new Cardiovascular event [10].

2.1.3.4 Rheumatism

This chapter will give a short background on the rheumatism pilot project, the project that this research is based on. It will discuss the reasons for the start of this project, as well as an elaboration on what has already been done so far and what the future steps are.

Background and incentives project

Rheumatism is an autoimmune disease that causes the body to attack its own joints and other parts of the body, causing inflammation and swollen, painful joints and surrounding areas [30] [15]. Further elaboration on the disease will be done in Chapter 2.2. For this project the choice has been made to focus solely on rheumatoid arthritis patients, which is a subcondition of rheumatism, since this is a chronical inflammatory disease with a fluctuating course, where disease flare-ups (exacerbations) can occur that want to be prevented. The risk of these flares occurring is especially high with decreasing the dosage of medication. The choice has been made to focus solely on RA and not on the other forms of arthritis because, although relevant, their disease-definitions, for example for flares, are less developed than with RA. Furthermore, decreasing the dosage is part of the treatment plan to a way lesser extent with these diseases than with RA. In addition to that, the flare-ups are less vehement in other diseases like for example normal arthritis as compared to RA.

The ultimate aim of this pilot project is to make sure that rheumatoid arthritis (RA) patients have a reduction of 15% in their medication use. Since RA flares occur very unpredictably, it is tough to make an estimation whether or not a patient can reduce their dosage. However being able to put such a reduction into practice would be a very desirable outcome, since RA medications carry a lot of side-effects, along with the fact that RA medication is very expensive [12]. Original medicines prescribed specifically for RA are also referred to as biologicals [16].

Therefore, the pilot project surrounds the development of two algorithms that both predict whether and in which time span a flare will occur, making decisions surrounding medication more responsible, better thought through and will ultimately result in less poli visits, higher efficiency of the specialists and easier reduction in medication or 'drug holiday'.

Consequently, the following question is to be solved during the pilot project:
Is it possible to predict with a dynamic model whether or not an individual RA patient, treated with biologicals, gets a flare of the disease? [10]

The results for this after one year has been as follows:
- predicting a flare with 78% accuracy with a joint modelling model that is being automatically updated with every visit.
- predicting a flare with 80% accuracy with a machine learning algorithm that is being automatically updated with every visit [10].

Contents project

As mentioned above, two algorithms have been developed to predict whether or not a flare will occur, and in which time frame. These are a joint modelling algorithm and a machine learning algorithm. Both algorithms have the same output, however both use different techniques to get to this result. The joint modelling algorithm has been developed by an in-house data scientist of the UMC Utrecht, who is now fine tuning the algorithm to see whether the above named percentage can be increased by taking various steps. Further information about the joint modelling algorithm is given in Chapter 2.3.

The machine learning algorithm has been developed by a data scientist who works at Siemens. Unfortunately, the contents or process of this development will not be included in this thesis, since Siemens has requested all their information to remain strictly private.

The reason that two algorithms were developed to predict the same outcome is because research has shown that joint modelling has proven to be an useful method to solve such problems, more information regarding the advantages of this is described Chapter 2.3.2. In addition, since one of the long term goals of this project is to expand the solution of this project in other diseases with a similar course, people want to know which method is best for solving such problems, so this approach can be used in a wider range.

Alongside the development of these two algorithms, a dashboard needs to be developed to put these results into practice, making them useful for both RA specialists and their patients. The development of a prototype of this dashboard is a part of the project that a part of my research will be focused on. More information about this can be found in Chapter 4.

Future steps

Future steps to be taken before this project is considered complete are the following: after both algorithms are completely finished, they need to be validated on a new external dataset. During this validation both models need to be evaluated, see how they are performing and it is tested which model is performing better. After that, a clinical trial needs to be done which includes the use of both models by select doctors with regard to rheumatoid arthritis patients. Based on that an evaluation is done on whether the model positively impacts the patient consults. If the answer to this is yes, then the model will be integrated in the daily use of the RA doctors, ideally with the use of a new dashboard. After this has been completed, steps can be taken with regard to the expansion of the model for diseases with a similar course.

2.1.4 Progress and future of the project

In Chapter 2.1.3, the first results of each of the pilot projects are shown. Some of the pilot projects are completed, such as neonatology, whereas others are still in the development phase. Planning wise, the ADAM project consists of three phases. These three phases are the pilot phase, the start in scaling up the project and the start in embedding the projects. As of now evaluations have been done and the decision has been made to enter the second phase, which is to start in scaling up the project.

2.2 Reumatoid arthritis

Rheumatoid arthritis (RA) is an autoimmune disease in which the body's immune system mistakes its own tissues as foreign and attacks them. The immune system attacks joints and other parts of the body [30], such as the skin, eyes, mouth, lungs, heart and blood vessels, liver, kidneys, blood and the nervous system [31].

RA is a chronic disease affecting over 1.3 million Americans and a total of 1% of the worldwide population. The disease affects women more than men, where women are up to three times more likely to develop rheumatoid arthritis than men are. Furthermore, women are more likely to develop the disease at a younger age than men. RA in generally starts to affect people between the ages of 30 and 60 years old, where on average people start to develop symptoms at age 60 [32].

This section will explain rheumatoid arthritis as a disease, what the symptoms and causes are, the treatment and RA flares and their attached measure called the DAS.

2.2.1 Symptoms and causes

Rheumatoid arthritis causes pain, swelling, and stiffness in the joints, and may cause severe joint damage, loss of function, and disability. The disease can last from months to a lifetime, and symptoms can improve or worsen over time [35]. Below the symptoms and the causes of the disease are extensively explained.

2.2.1.1 Symptoms

Rheumatoid arthritis can appear in the upper limbs, the lower limbs and the spine and axial joints, with symptoms including:

- **Inflammation:** joint swelling with characteristic soft tissue involvement, meaning the swelling of the soft tissue of the body such as muscles, blood vessels, fat and so forth. The joint is also pink or red and feels very warm.
- **Swelling:** when the immune system attacks the lining of the joints, fluid then builds up around the joints that causes the swelling [34].
- **Polyarthritis:** the involvement of more than five joints.
- **Symmetry:** same joint regions of both upper and lower extremities.
- **Erosions/erosive damage:** erosion is a loss of bone from a disease process. With rheumatoid arthritis, bony destruction can be seen on an conventional x-ray.
- **Painful metacarpophalangeal (MCP) or metatarsophalangeal (MTP) compression:** metacarpophalangeal joints are the largest joints of the hand, also referred to as knuckles. RA patients often have these joints inflamed and painful. Metatarsophalangeal joints are the joints between the metatarsal bones of the foot

and the proximal bones of the toes. Inflammation, pain and swelling of these joints are very common with RA.

- **Morning stiffness (lasting more than a few minutes):** morning stiffness and therefore pain in the joints after long periods of inactivity, like sleeping [33] [35].

The progression of the disease can be distinguished in four phases:

1. **Initial phase:** some patients may have markers in the blood that denote autoimmunity, this is an early inflammatory phase that includes clinical manifestations that may or may not be diagnosed as having RA.

2. **Destructive phase:** this phase includes erosions, so the destruction of bone tissue, and disease progression.

3. **Third phase:** an ongoing phase that involves irreversible joint construction.

4. **Fourth phase:** presence of either the rheumatoid factor (RF) and/or presence of antibodies that can bind cyclic citrullinated peptides (CCP), which is an auto-antibody produced by the patients immune system that attacks the body [32]. Patients with either these of these biomarkers present in their blood tend to have a more severe course of RA, with anti-CCP antibodies having a greater prognostic value than RF. Patients with neither of these biomarkers present in their blood tend to have a more gentle course and are referred to as having "seronegative" RA [35].

2.2.1.2 Causes

What happens in the body of an RA patient that causes inflamed joints?

When a person suffers from an autoimmune disease, his or her immune system mistakenly attacks healthy tissue. Specifically within RA, this healthy tissue that the body attacks is the tissue that is situated within the joints. As can be seen in figure 2.3, the synovium is the membrane that lines a joint or surrounds a tendon and releases fluid, allowing for joint movement. In people suffering from RA, white blood cells create inflammation in this synovium, causing the tissue that covers the walls of the joints to thicken and become swollen and painful when moved. This inflammation often leads to joint erosion, a loss of motion and joint damage. Overtime, the affected synovium destroys the cartilage and bone within joints, which causes everything around that area that is supposed to support the joint, like muscles, ligaments and tendons, to weaken [32].

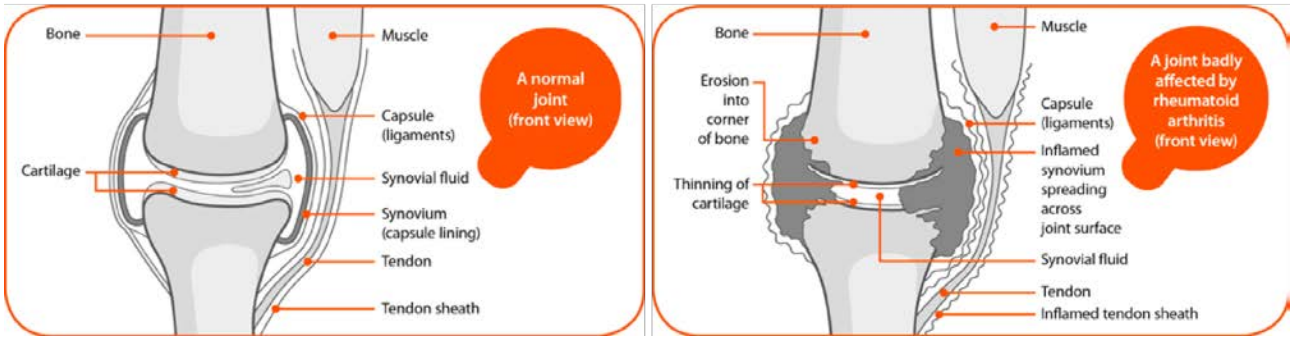

Figure 2.3: Representation of a joint affected by rheumatoid arthritis. [36]

Causes of the presence of RA within a person:

Below the common known causes of the development of the disease within patients are explained. Research about this is still being done to this very day.

- **Genetics:** approximately 60% of RA susceptibility is attributable to heritability [35].
- **Environmental sources:** smoking is one of the biggest environmental factors associated with developing RA. Smoking increases the risk of developing seropositive, which means that the rheumatoid factor blood test is positive. The risk of developing seropositive RA is dose-dependent and increases with the number of years ever smoked. Former smokers remain at risk for RA for anywhere between 10 and 19 years after quitting smoking. Other environmental factors contributing to increasing the chance of developing RA include living in regions with greater air pollution, as well as women who had a higher birth weight, i.e. larger than 4.54 kg [35].
- **Interaction of genes and environment:** a presence of a certain type of genes, namely PTPN22 genes, within a person coupled with smoking contributes to heightened RA risk [35].

2.2.2 Treatment and medication

The primary goals in treating patients with rheumatoid arthritis are to reduce pain and stiffness, slow disease progression (prevent damage), and improve function. There are several medications available to treat this, like non-biologic and biologic Disease-Modifying Antirheumatic Drugs (DMARDs) that make pain relief and the delay of disease progression a realistic possibility, as well as nosterodial anti-inflammatory drugs (NSAIDs) and corticosteroids to minimize the toxicities. Most of the times, a combination of these medications are used, however the treatment differs per patient [35].

For this research, as mentioned before, there is a focus on RA patients who make use of biologicals. Biologicals are genetically engineered proteints that are made from human genes. Unlike other RA medications that affect the entire immune system, biologicals specify on the parts that control the inflammation process. Biologicals have been shown to slow or even stop joint damage from getting worse [39].

The UMC Utrecht makes use of the following biologicals, that have the following ATC codes in the data:

- **L01XC02:** Mabthera
- **L04AA24**: Orencia
- **L04AB01:** Enbrel
- **L04AB02:** Remicade
- **L04AB04:** Humira
- **L04AB06:** Simponi
- **L04AC07:** Roactemra

2.2.3 DAS

The Disease Activity Score, or DAS, is the generalized metric that is being used to measure the activity of patient's rheumatoid arthritis [37]. A modified version of this is the DAS28, which focuses on the assessment of 28 joints instead of 44 joints. The DAS28 is a formula that consists of the following components:

- **Tender28:** the number of tender joints, meaning the number of joints that are painful. There may be pain when you press or when you move the joint [41]
- **Swollen28:** the number of swollen joints, meaning the number of joints that look or feel swollen [41]
- **CRP:** the C-Reactive Protein level, which is a substance produced by the liver in response to inflammation [42]
- **GH:** the patient global health assessment (from 0=best to 100=worst), so how is the patient feeling. [40].

These components are then combined in the following formula:
$\text{DAS28} = 0.56 * \sqrt{\text{tender28}} + 0.28 * \sqrt{\text{swollen28}} + 0.70 * \ln(\text{CRP}) + 0.014 * \text{GH}$

DAS28 provides a number on a scale from 0 to 10, with values of > 5.1, < 3.2 and 2.6, indicating high disease activity, low disease activity and remission, respectively [37]. These measurements have been further altered by the UMC Utrecht, resulting in the following criteria for a flare:

- Increase in DAS28 > 1.2 with ≥ 1 increase in SJC and DAS28 ≥ 2.6 OR
- Increase in DAS28 > 0.6 and DAS28 ≥ 3.2 with an increase of SJC ≥ 1 OR
- Increase in medication (dose category) without DAS28 < 2.6 (either missing or DAS28 ≥ 2.6 [16].

2.2.4 Flares

As mentioned before, flares are an increase in rheumatoid arthritis symptoms. There are common factors that have been proven to increase the chance of having a flare. These are as follows:

- **Changing the dosage of the medication or switching medications.** About 60% of RA patients had a flare when tapering off of their medication [17].
- **Getting an infection or virus**. When an infection or virus enters the body, the immune system ramps up its activity and therefore it can induce or trigger an autoimmune disease [18] [19].
- **Smoking or being around smokers for a long period of time**. Research shows that current smokers had higher flare rates than non-smoker, and former smokers were not different from non-smokers. This shows that smoking therefore adversely affects RA flares [20]. Aside from that, rheumatoid arthritis smokers are three times as likely to have the rheumatoid factor, which is a sign of a more severe disease, and twice as likely to have joint damage [21].
- **Weight gain**. Fat contains inflammatory chemicals that impact the muscoloskeletal and cardiovascular systems, including adipokines, which is a type of cytokine that is released by fat tissue. Adipokines are known for promoting inflammation, which therefore contributes to the intensifying of flares. In addition to that, people who suffer from obesity have an increase in load on the joints, which will make the pain worse [22].
- **No exercise**. As explained above, excess weight contributes to an increase in flare intensity. Aside from that, exercise is shown to be a factor in controlling inflammation [23]. People suffering from rheumatoid arthritis who do not exercise cannot benefit from these advantages, resulting in a higher susceptibility for flares compared to rheumatoid arthritis patients who do exercise.
- **Stress**. Stress has been known to affect the occurrence of rheumatoid arthritis flares for decades [24]. Stress in general has got a negative impact on the body, since it triggers a cascade of stress hormones that produce well-orchestrated physiological changes. Examples of this are increasing heart rates, quickening in breathing and muscle tensing [25]. The hormones that are released during stress are known for their stimulative effect on immune cells, making stress worse for people suffering from rheumatoid arthritis. They may therefore in turn activate key immune mechanisms associated with greater inflammatory response, leading to a disease flare [24].
- **Weather**. It has been shown that in most patients, weather changes increase arthritic symptoms. Women are more sensitive to weather than men. The weather factors that appear to have the most influence over RA symptoms are high barometric pressure, low temperature and high relative humidity [26] [27] [28].
- **Hormones and Pregnancy**. Estrogen and progesterone have found to offer protective properties against rheumatoid arthritis, which is possibly why about three times as many women suffer from the disease than men. Over time female sex hormones decrease, which is why majority of women diagnosed with rheumatoid arthritis are over the age of 45. Aside from that, research has shown that women's symptoms

improve during the postovulatory stage of their menstrual cycle and during pregnancy, but worsen after childbirth and during the second week of their menstrual cycle [29].

2.3 Joint modelling

One algorithm developed to predict flares in rheumatoid arthritis patients, has been developed making use of the joint modelling method. Joint modelling is a method that allows for simultaneous analysis of both longitudinal measurement outcomes or time-series data, and time-to-event outcomes, also called survival data [43]. In this section the joint modelling technique will be discussed, along with the relevance of this technique to problems such as these and how this technique has been implemented in this particular problem.

2.3.1 Definition of joint modelling

The hospital has been collecting two types of data in the RA field, namely longitudinal data and time-to-event data. The longitudinal data collected by the hospital for this problem result from regular doctors visits and blood tests, which includes number of swollen or tender joints, blood pressure, age and so forth. An expanded overview of all the relevant longitudinal variables collected can be found in Chapter 3. The survival data was later derived by the data scientists of the hospital. Thus, the collection of data is a continuous process.

Most of the time, especially in questions such as these, associations and dependencies can be drawn between these two types of data, as a change in longitudinal data often accounts for an event to happen. Joint Models for longitudinal and survival/time-to-event data are models that bring these two data types together into a single model so that the dependence and association between the longitudinal biomarker, which is a process that can be measured from data of the patient and predict the outcome of a flare, and the time to event can be seen to better asses the effect of a treatment [45].

Joint models consist of two parts, namely a model for event occurrence and a model for trajectory of longitudinal measurements that share some parameters. These two parts will then be linked.

2.3.2 Reason choice of technique

This technique has been chosen by the UMC Utrecht because of two reasons:

1. There was a wish for another algorithm, developed making use of a different technique than machine learning, so that both algorithms can be compared and it can be researched which technique, when successful, will be expanded for solving similar problems

2. Joint modelling algorithms have proved to be very useful in the past to solve similar problems. Examples of problems that are solved with this technique are predicting abdominal aortic aneurysm growth and rupture, application to liver transplantation data and application to psychosis data [49] [50] [51].

When questions such are these are posed, which is to predict whether and when a flare will occur, the focus lies on the time-to-event process. To solve these processes, so-called naive approaches with a time varying co-variate have historically been applied. Naive approaches are estimating techniques in which the last period's actuals are used as this period's prediction, without adjusting them or attempting to establish causal factors. However, a great number of studies have demonstrated that it is preferable to model both outcomes jointly, in particular when the censoring mechanism is informative of the time varying covariate is measured with error, as efficiency is increased and bias decreased with this method [48]. Joint models provide more efficient estimates of the treatment effects on the time to event, as well as the longitudinal marker and reduce bias in the estimates of the overall treatment effect [45].

Because of the above advantages, Joint Models are increasingly used in clinical trials, since a less biased estimate leads to a more accurate estimate of the treatment effect [45]. However, joint modelling does come with its disadvantages. These are the increase in computational effort required to fit the models and the relative scarcity of software to enable their routine use [46].

2.3.3 Joint modelling algorithm for predicting flares

The joint modelling algorithm for predicting rheumatoid arthritis flares, is being developed in R, a programming language, by an in-house data scientist of the UMC Utrecht. The joint modelling algorithm works as follows: from the dataset a subset is made where the longitudinal marker, meaning the target, is always present, along with the course, which is the first moment when a patient starts to reduce in medication up until a patient gets a flare. Based on this subset by comparison it the whole dataset, the appropriate courses are determined. Then the patients who experienced a flare are distinguished from those who did not. Based on the above, the first original joint model is defined with one group and the model is trained with the use of cross-validation. Then dynamic predictions are done on the test set with a fixed horizon, which is the period you predict in advance, of 12 weeks and variable landmarks, which are the points from which the predictions start. These time points are then matched to the actual ones in the data, so that the predictions can be compared. A more detailed explanation of how the algorithm works can be found in Appendix A.

2.3.3.1 Progress joint modelling algorithm so far

As mentioned in Chapter 2.1.3.4, the joint modelling algorithm is now generating 78% accurate predictions. Activities that are done now to fine tune this algorithm is to develop a good system to be able to do nested cross-validation. After this is done, the model is complete and can then be fitted and tested properly on a newly generated dataset.

2.4 Machine learning techniques and tools

The tools used for the comparison of the joint modelling algorithm with standard machine learning algorithms, the programming language Python 3.6.3 is used in combination with Anaconda. The data investigation is done with Pandas, Excel and SAS. The pre-processing and feature construction part is done with Pandas, visualizations are done with Seaborn and Matplotlib and the machine learning algorithm developed for the validation of the joint modelling algorithm is done making use of Scikit-Learn.

2.4.1 Scikit-learn implementation

In this thesis, supervised learning is used to predict flares, since we are dealing with labelled data [73]. Supervised learning is a machine learning task that maps an input to an output based on example input-output pairs [52]. There are many possibilities within supervised learning, like neural networks, logistic regression or decision trees. In this thesis, the choice has been made to make use of classification techniques, in particular the Decision Tree Classifier (DT), the Support Vector Machine (SVM) and the Stochastic Gradient Descent classifier (SGD). Classification is a supervised learning approach in which the model learns from the input data and then uses this to predict the classification of a new observation. A choice has been made to make use of classification since we are predicting a category, namely whether or not a flare will occur. Furthermore, classification techniques, in particular decision tree, give interpretable models, something that is very useful in the comparison of standard machine learning algorithms to the joint modelling approach, since this allows for an eased comparison between the two.

The decision tree method builds a classification model in the form of a tree structure. It breaks down a data set into smaller subsets while incrementally developing the associated decision tree. The final result is a tree with decision nodes and leaf nodes. Leaf nodes represent a classification or decision, so in this case whether a flare will occur or not. Decision nodes have two or more branches. The topmost decision node in the tree, also called the root, corresponds to the best predictor [53]. The decision tree classifier used is processed in the tree package in scikit-learn.

A support vector machine is a selective classifier which results a separating hyperplane. When given labeled training data, the algorithm generates an optimal hyperplane which categorizes new examples as output [54]. This support vector machine classifier is processed in the svm package in scikit-learn.

The stochastic gradient descent classifier implements regularized linear models with stochastic gradient descent. This SGD considers only 1 random point while changing weights instead of considering the whole training data [55]. This SGD is processed in the linear_model package in scikit-learn.

Chapter 3

Data

This chapter will give an outline of the dataset that is used in the research conducted. In Chapter 3.1 a short background and the sources of the dataset will be given. Secondly in Chapter 3.2 general information about the data will be explained, along with an elaboration on the target variable. In Chapter 3.3 relevant variables with matching descriptive statistics will be discussed and finally in Chapter 3.4 the missing data of this dataset will be elaborated on.

3.1 Background and sources

The data that is being used for this pilot mostly comes from HiX, which is the electronic patient file of the UMC Utrecht the platform on which doctors/specialists record their patient data, but also from the medical registration of the pharmacy. To enclose this data in a secure way, a platform is created that consists of copies of diverse data tables of different departments. This platform is called the Research Data Platform (RDP), and this is also the location where the editing and pre-processing of the data that will be used by the models, will take place.

There are three sources that together form the entire dataset called the Analytical Base Table (ABT), the eventual table that is being used to model on [16]. These sources are as follows:

1. **MED1910:** The medication data. Source: RDP and the pharmacy
2. **DAS2610:** lab and other clinical (disease activity) measurements as well as patient demographic data. Source: RDP
3. **Other data:** like smoking, erosion, first diagnosis dates, etc. Source: RDP [16].

More on the process of how this ABT is constructed can be found in Appendix B.

3.2 General information

The type of data that is available is time-to-event data or survival data and time-series data or longitudinal data. The time-series data is the data of the rheumatoid arthritis patients that has been collected regularly at doctors visits that occur roughly every three months per patient over the past years, with the earliest measurement dating to 2012.

The data that has been collected during doctors visits as time-series data are as follows: a series of RA related blood values, swollen joint count (SJC), tender joint count (TJC), VAS, whether a patient smokes or not (Smoke), age and medication information such as Dosage and Name of medication.

The time-to-event data is, as mentioned in Chapter 2.3, the recorded time and data until a flare happens to a patient. This is represented with the variable ID, where one ID is the data of the course of a patient until a flare happens. When this flare happens, the course of the patient is considered 'dead' and a new ID is assigned to represent a new course. The variable New_Event is the variable that states the occurrence of a new flare after there has been tapered with a patient, meaning a reduction in medication has taken place.

The properties of the ABT are as follows: the dataset consists of 8542 entries or rows, 75 variables or columns and there is data present of 303 distinct patients. This dataset includes the contents of the three sources mentioned above, as well as helping variables added to ease the modelling process. More on these variables can be found in Appendix B.2.

The number of entries present per patient are as follows:

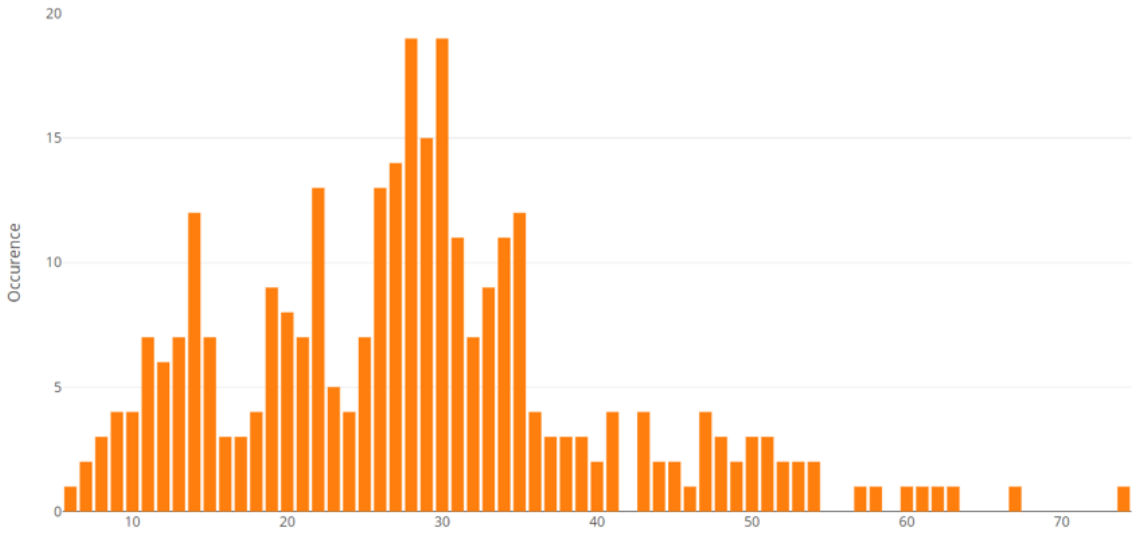

Figure 3.1: Number of entries present per patient.

3.2.1 Target

As described in Chapter 2.2.3, the DAS value is a weighted sum of the following four components:

- Tender28: the number of tender joints
- Swollen28: the number of swollen joints
- CRP: the C-Reactive Protein level
- GH: the patient global health assessment (from 0=best to 100=worst)

These components are named the following in the dataset:

- Tender28: TJC
- Swollen28: SJC
- CRP: CRP
- GH: VAS

The target variable in the dataset for the joint modelling algorithm is New_Event. The focus is solely on patients having flares after tapering, as mentioned in Chapter 2.1.3.4. This New_Event is a boolean variable is based on the DAS28EST variable, and turns 1 when this variable meets flare criteria according to the requirements set in Chapter 2.2.3. DAS28EST is a mean of six DAS28 variables, each slightly differing the weight of the value in the blood (either CRP or BSE, where BSE is the Erythrocyte sedimentation rate, a measure of inflammation) and the VAS value. More on the exact calculation of the DAS28 is described in Appendix B.3.

Other target variables worth to consider for the development of the machine learning algorithm later on in this research are New_Flare, which is the occurrence of a flare without a patient having tapered and (all six versions of) the DAS28 variable.

3.3 Relevant variables

This chapter looks at descriptive statistics of variables present in the data that could be useful in the development of the machine learning algorithm. These variables include Age, DAS, Weight/BMI, tapering, New_Event, Smoking/non-smoking and medication.

The average age of the people in the database is 50.37 years, and the age distribution is as shown in Figure 3.2. Although there is a decent distribution over almost all age categories, the peak in patients is roughly from 40 years onward.

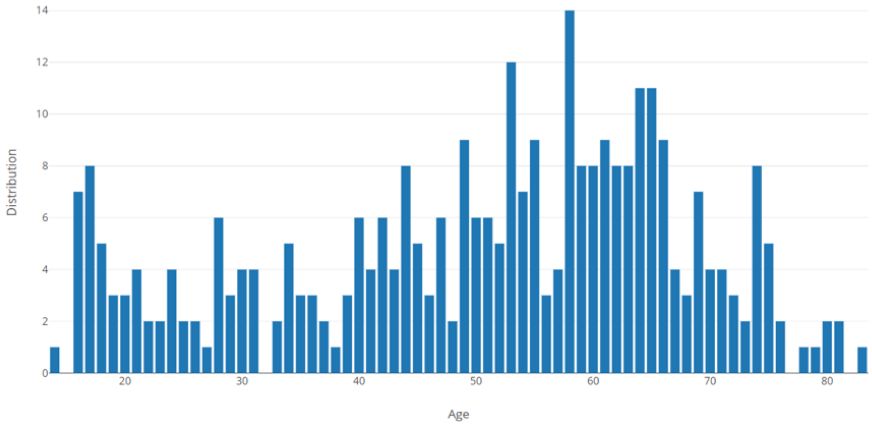

Figure 3.2: Age distribution of patients.

The distribution of the DAS values recorded is shown in Figure 3.3. The average DAS value is 2.45. Furthermore, Figure 3.4 shows a distribution of the number of DAS values present per patient.

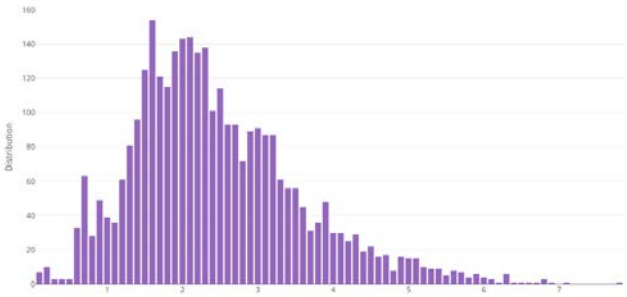

Figure 3.3: Distribution of DAS values.

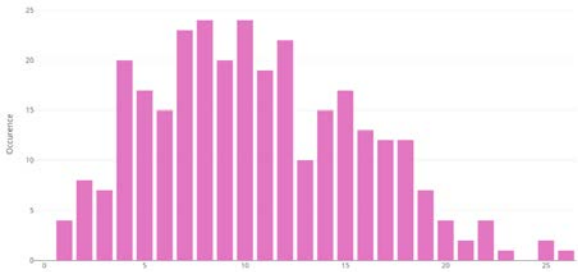

Figure 3.4: Number of DAS values present per patient.

As discussed in Chapter 2.2.4, weight turns out to have a significant effect on rheumatoid arthritis symptoms. However, a change in weight has not been recorded during visits. Rather, the weight has probably been recorded once at the intake consult and therefore remained constant in the database throughout the course of the treatment. Regardless, based on the data an overview has been made in Figure 3.5 based on the amount of flares occurring per BMI group. As can be seen in this figure, the amount of flares and events increases significantly as the weight increases. The average flare and event count per patient respectively was 0.81 and 1.22 with those who classified as underweight, 2.72 and 2.01 with those who classified as having a normal weight, 3.18 and 2.30 with those who classified as overweight and 3.85 and 2.71 with those who classified as obese. As can be drawn from these figures, weight has a clear and significant impact on the amount of flares/events occurring.

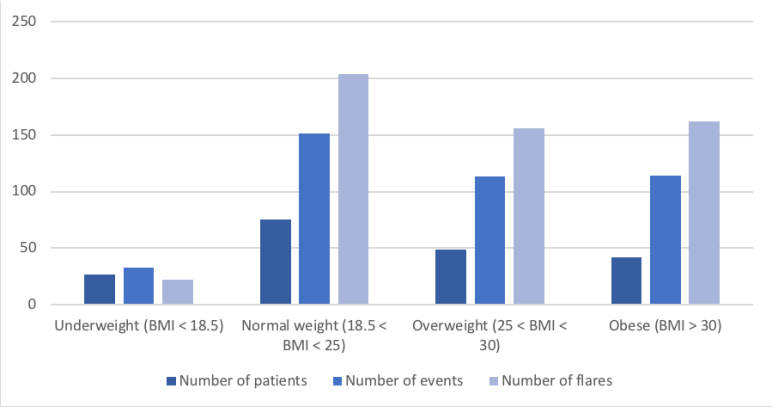

Figure 3.5: Distribution of flares per weight class.

The number of RA patients that have been or are currently tapering, is 291 of the 303. Within this number of patients, 134 new events (flares) have occurred. Figure 3.6 shows a distribution of in which week of the treatment a flare (after tapering) occurs.

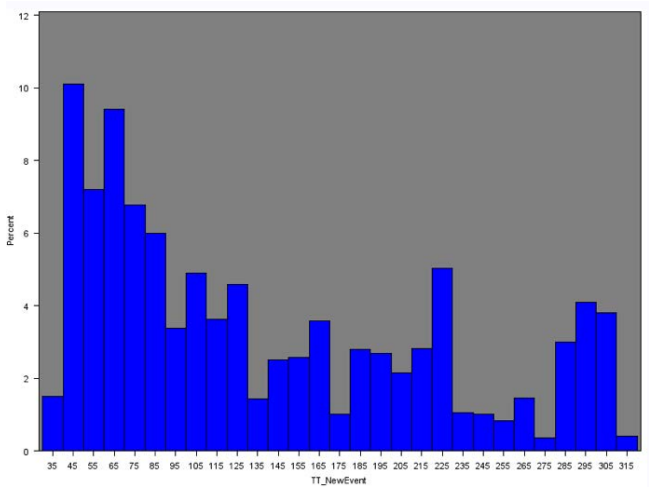

Figure 3.6: Week of treatment in which a new event (flare) occurs.

The amount of patients smoking has also not been recorded in a big way, as will be explained in the following chapter. There are 1351 entries of smokers, so 46 patients and 1890 entries of nonsmokers, so 70 patients. There is one recording of a transition between the two. On average, 1.9 flares occurred per smoking patient before tapering, and 2.4 flares occurred after tapering per smoking patient. Compared to non-smokers, an inverse result was shown to what we expected, since on average 3.0 flares occurred per non-smoking patient before tapering, and 2.8 flares occurred after tapering per non-smoking patient. However, a lot of this data is missing, as discussed in Chapter 3.4, along with the fact that the data has not been updated in between.

There are 7 different biologicals prescribed for RA patients. More info on these biologicals can be found in Chapter 2.2.2. The medication distribution amongst the patients is as follows, accompanied by the number of flares and events per medication:

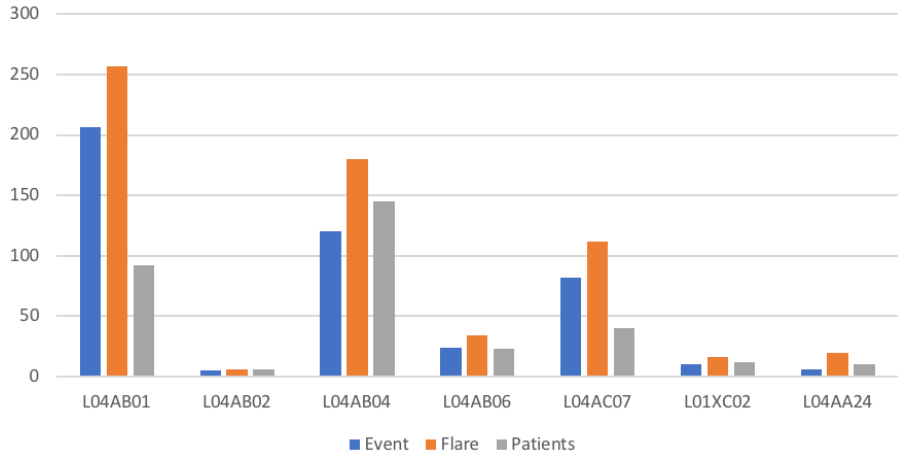

Figure 3.7: The distribution of medication across patients.

3.4 Missing data

Unfortunately, although a lot of data has been collected over the past years, a lot of it is missing. Of the 8542 entries, in 5335 entries the DAS was missing. Aside from that, a lot more variables had a significant chunk of missing values. Appendix B.4 contains a table with all the variables that has the most missing data. Below the most essential variables of these are highlighted.

The four components referred to in the above Chapter 3.2.1, are essential to calculate the target variable DAS28. However, investigation into the percentages of data present has shown that all four variables are missing a lot of data. The percentage of missing data of each of the four variables are as follows: TJC is missing in 59.4% of the cases, SJC is missing in 58.7% of the cases, the VAS is missing in 68,1% of the cases and the CRP is missing in 77.6% of the cases. This results in a very high percentage of the DAS28 variable to be missing, namely 84,6% (see Appendix B.4).

Other useful variables for predictions that are missing a lot in the data are the DAS28Date, which is the date on which the DAS28 has been recorded, BSE, BSElab, which are certain blood values used in the place of the CRP value to calculate an alternative DAS28. Others are BSEDate, CRPDate, VASDate, which are the dates on which BSE, CRP and VAS have been recorded, and Smoke and Erosive, which records whether or not a patient has erosion occurring in their joints. These variables have not been recorded enough, and so the effect of these variables, that might have an effect on the outcome as discussed in Chapter 2.2.4, can not be determined at this moment with the use of algorithms.

Smoke, DAS28Date and VAS28Date, as well as SJC, TJC and VAS can be measured and recorded during doctors visits, whereas BSE, BSElab, BSEDate, CRPDate and Erosive can only be found and therefore recorded with the use of blood tests.

Chapter 4

Exploratory Dashboard Visualization

This chapter discusses the development of a dashboard prototype designed especially for rheumatoid arthritis specialists and patients. The outline of this chapter is as follows: first in Chapter 4.1 the reasons as to why this dashboard should be developed will be considered, both from a hospital/specialist point of view as well as a data visualization point of view. After that, the scope of the dashboard will be clearly defined in Chapter 4.2. Chapter 4.3 shows the dashboard as it is being used by the RA specialists at this moment. Chapter 4.4 discusses relevant usability research and principles relevant to the development of the prototype. Chapter 4.5 will elaborate on the development process of the dashboard, taking into account all the steps taken and reasoning behind them. Finally in Chapter 4.6, the future steps with regard to putting the dashboard into practice will be discussed.

4.1 Reasons for development

There have been many motives for the development of a dashboard that will be used by RA specialists and patients during consults. Below these reasons, for both RA specialists and the hospital as well as the data scientists working on this project, are discussed.

4.1.1 Hospital and specialists

Reasons for development include first and foremost that the results of the machine learning and joint modeling algorithms should be represented in an insightful manner to the specialists, so that they can use this in their daily consults to improve treatment. Secondly, as will be elaborated on in Chapter 4.3, the current dashboard has a lot of options that are scattered all over the place, resulting in high inefficiency among the specialists. Since consults for patients suffering from RA is done in a routinely manner, there was a wish for a dashboard that included all tasks undertaken during such a consult, presented in a structured and clear way, where usability and efficiency is highest priority.

Aside from what is stated above, improved efficiency is also achieved by designing a dashboard that reduces cognitive overload, which, when achieved, creates a lot of benefits for the specialists. Even though over the past years there have been massive improvements in display technology, the quality and clarity of visually represented clinical data in these displays continues to be of low quality. These constraints add to ongoing cognitive overload for hospital specialists, increasing the chance for diagnostic error. Human factor studies have demonstrated that 80% of "user error" is attributable to cognitive overload [56]. A dashboard designed to reduce cognitive overload and improve operational efficiency is therefore of paramount importance to RA specialists and the hospital itself.

4.1.2 Data scientists/developers

Aside from the benefits that the dashboard brings to the RA specialists and the hospital itself, a successfully implemented dashboard also brings several benefits to the data scientists working on the flare prediction algorithms. As discussed in Chapter 3.4, a large percentage of the necessary variables, such as the DAS and its components, is missing. This is is because of several reasons, mostly caused by the complexity of the current dashboard, as can be seen in Chapter 4.3. This results, because of time constraints, in less recording of data, or putting the relevant data in the wrong place. Missing data complicates and hinders the modeling/coding process, since less reliable data is available to make predictions on. Therefore the accuracy of the developed algorithms will likely be lower than algorithms developed using more complete data. Aside from the obvious necessary variables like the DAS and its components, other data as described in Chapter 2.2.4, such as smoking and other factors that influence RA flares, could have a significant effect on the model and help to predict the target variable.

Since this dashboard will clearly and structurally display easy to use boxes for all the necessary data, this will contribute to a larger collection of this necessary data. Furthermore, since one of the main elements of the dashboard is a graph of the possibility for flares with patients, this will further encourage the RA specialists to record the necessary data since they are aware of the fact that this will help improve the graph, and therefore provides them with more information that could help their patients.

4.2 Scope

The scope of this part of the research is as follows: based on the relevant research and usability methodologies, the requirements of the specialists and data scientists and the focus group, a prototype of the design of the dashboard should be made.

This dashboard will only be developed for consults surrounding patients suffering from RA. Since these consults occur roughly every three months and the same topics are covered, this is routine work that could be simplified by the use of a specific dashboard. Therefore the different topics treated during such consults should

be included in the dashboard, as well as the results of the algorithms for the particular patient. Furthermore, the components necessary for the data scientists should be included this design.

4.3 Current dashboard

The dashboard that is being used at the moment is from HiX, a system used in the hospital for the registration of patient information. Even though HiX helps digitization within the hospital and has all necessary information available, it is very unclear and many different features are scattered all over the place, and, above all, it is not very user-oriented. This is illustrated by the fact that there is a large number of tabs present. There are a lot of separate tabs present for tasks that could easily be merged, such as the cover page, consult notes and the joint scores, since these are elements that are done during the first phase of the consult. Furthermore, if the specialists do not construct a personalized favorites list that is displayed under another tab in the dashboard, the different tabs and functionalities necessary to perform their jobs well are not structured in any way. In addition to that, the specialists are faced with a lot of unstructured information that is presented in each of these tabs, resulting in cognitive overload. In Appendix D.1 the HiX Electronic Patient Document is shown, which is the dashboard that the rheumatism specialists currently work in.

4.4 Usability research and principles

As mentioned before in section 4.1.1, the aim is to create a dashboard that is user-friendly and reduces cognitive overload, reducing decision-making error. Cognitive overload is the situation when the required amount of information to complete a task (in this case to understand a piece of information presented during the task) is too much to be processed easily in working memory [58]. Numeric and textual data analysis by doctors and specialists results in excessive cognitive strain and irregular thinking patterns, all of which impact the quality of care and patient safety [56].

Therefore the dashboard was designed to spatially and in a human-centered way organize RA patient data, reducing cognitive overload. One way is a reduction in the number of clicks, taking into account the flow of attention, simplicity in design, the most important items should be on the first page, the dashboard should be a helping tool in guiding the user through the process in which tasks are completed and making use of grouping, color and visuals. The first page of a dashboard should include the most important tasks that will be undertaken. Aside from that, it should reveal the rest of the content that the dashboard has to offer, along with easy access to other relevant features [65]. The number of clicks illustrates task efficiency, where a high number of clicks illustrates a low task efficiency and a low number of clicks illustrates a high task efficiency [57]. Therefore, the less clicks used in any dashboard, the more efficient the dashboard is.

Another way to reduce cognitive overload and improve efficiency is to design a dashboard that takes human focus, in particular the flow of where the eyes go, in a digital environment into account, according to the

research done by the Norman Nielsen Group. Users of a web-environment read in a F-shape, starting with horizontal moves on the top headlines and fast vertical scanning of the body and sub-headlines [58] [59].

Another integral component of increased human-centered design and reduction of cognitive overload is simplicity in design. Complexity is often perceived as being the enemy of usability [60]. Computer information systems implemented in healthcare environments can often become agonizing and inconvenient, turning an opportunity for improving care into a potential cause of errors and confusion [61] [62]. This is why simplicity in design is of paramount importance. System complexity increases as the number of features increase. Therefore, the number of features must be minimized to pursue simplicity [63] [64]. To minimize the number of features present, the design should start with the users, asking the question with every feature: "is this information truly useful to users?" [70].

One of the main outcomes of a usability interface is that it should be easy for users to achieve their objectives through using the dashboard. A good design will guide them through that process [66]. Therefore, the design of the process must be structured and following the order in which the tasks are executed.

Grouping similar tasks in the same interaction scenario enhances usability and with that efficiency and a reduction in cognitive overload [67]. Rightful grouping of tasks creates overview and structure in the brains of the users. According to the principle of Gestalt, proximity means that when objects are placed together, the eye perceives them as a group [67]. This reduces visual clutter and creates an overview, making the user more inclined to perform this set of tasks together.

Use of color has proven to have big effects on how users perceive certain information [68]. Colors like red, orange and green are said to be actionable colors, since they resemble traffic lights. These colors can therefore be used as warnings or severity of a situation, building up from green to orange to red. The color red is associated with warnings and importance. Orange has two existing associations: that of immediate instructions to carry out, and that of waiting. Green is associated with success and encouragement. Therefore, making use of colors and these colors in particular, will have an impact as to how the user will perceive information [69]. Aside from this, several shades of the same color can draw importance to that particular element. For example, when a lighter shade of a color (less opacity) is used, this draws less attention to that particular element, whereas when a higher intensity of this color is used, it will draw more attention to that particular element [70].

Lastly, fitting use of visuals, when chosen carefully, can aid users in a significant way. Visuals can help users to interpret relevant data, especially if they are visuals the users are familiar with [70]. However, visuals should not be overused since only a limited amount of cognitive processing can take place in the visual channel at any one time [71].

4.5 The development process and results

Below the development process and results of the dashboard prototype are discussed. This is done in the following steps: first, an introductory discussion with the leader and data manager of the project to find out

what the requirements from their side are for the dashboard (as described in Chapter 4.2). This was followed by an email that was sent to one of the RA specialists who helped me during this research. The email included a couple of preliminary requirements engineering questions. This was followed by an observatory visit to the rheumatism polyclinic, based on which a first version of the prototype was developed. This prototype was evaluated along with some usability questions with the RA specialist, and adjustments required were processed in prototype version number two.

4.5.1 Preliminary gathering of information

As a part of the gathering of information several initial requirements were gained with the aim to get a first insight into the strong and weak points of the dashboard. Conclusions that came out of this were as follows: currently, a lot of clicks are needed in the EPD during the consult which is highly inefficient. Furthermore an automatic DAS calculator is preferable, along with a pop-up of whether a patient is eligible for research and a nicer visual of the disease course. The information that is required to be present at first glance is the diagnosis, history of the patient, medication history and use and the information of last consult. The full email with the concerning questions and their answers can be found in Appendix D.2.

4.5.2 Polyclinic visit and interview

During the polyclinic visits insights were gained as to how the dashboard is being used, what goes well and what the bottlenecks are. Many patients with different backgrounds (rheumatism related) visited these consults, most of them that day were suffering from rheumatoid arthritis. My observations from seeing these consults are as follows: during the consults a lot of to-and-fro clicking between the different tabs occurred. The doctor who offered me the opportunity to observe her during her polyclinic visits explained to me that she had constructed a favorites list of all the tabs that she used most often, so that she would not get as lost in all the different tabs present. This list can be found in Appendix D.3. The results were scattered all over different tabs, including the different necessary variables, as can be seen in Appendix D.3. The same tasks were used for RA patients coming for a regular check-up consult. The order in which the tasks of the consult progressed for RA patients can be found in Appendix D.4.1.

Contributing factors to this to-and-fro clicking were the fact that all relevant results all had their separate tab, as well as all the tasks that needed to be executed were scattered all over different tabs, as can be seen in Chapter 4.3.

Furthermore, a lot of terms needed to be entered manually, such as a list of questions she repeatedly asked with every consult, which added together, results in a lot of inefficiency. The rating of the VAS with the patient also is an area where time can be won because of the representation within the dashboard. As shown in Chapter 4.3, it is a number that needs to be filled in, but there is no indication of what is best and what is worst, which is often confusing to the patients. In addition to this, the overview of medicine use was not clear, what precisely is being used and therefore what prescription needs to be extended. Furthermore, the

list of medicines was not alphabetically ordered, or ordered in categories like for example biologicals, pain medication, and so forth, which caused the doctor to spend a significant chunk of her time trying to locate the exact medicine needed. The complete list of observations from the visit at the polyclinic can be found in Appendix D.4.

After this visit at the polyclinic, further questions needed for the development were discussed. This can be found in Appendix D.5.

4.5.3 First Prototype

Based on the information gathered from the above, a first prototype was developed. This prototype looks as follows:

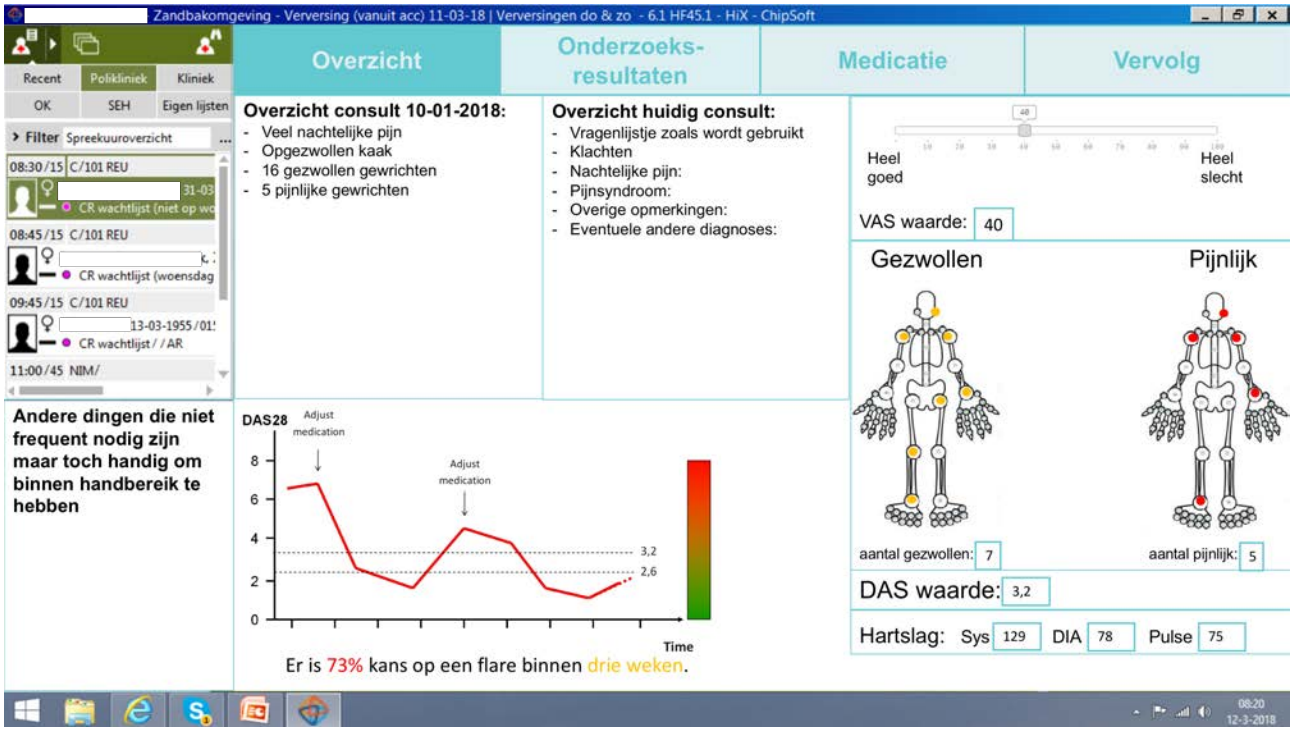

Figure 4.1: Dashboard page 1

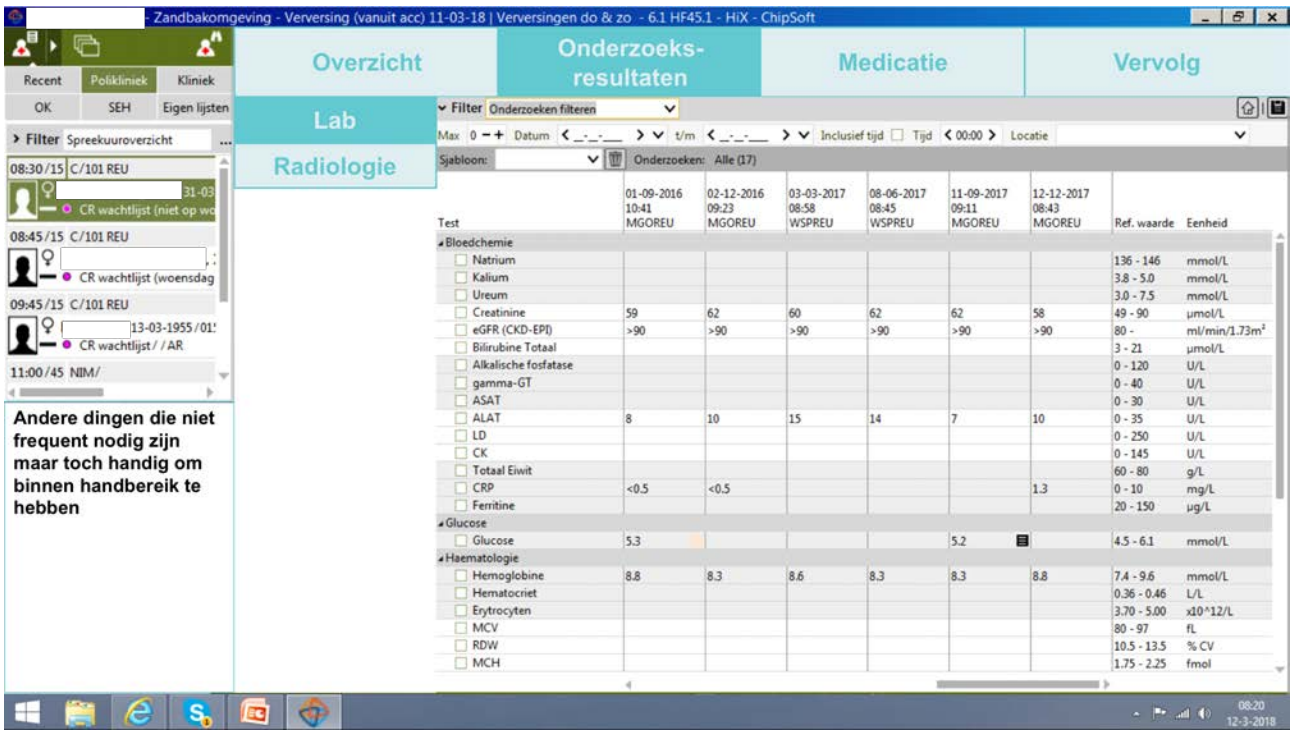

Figure 4.2: Dashboard page 2

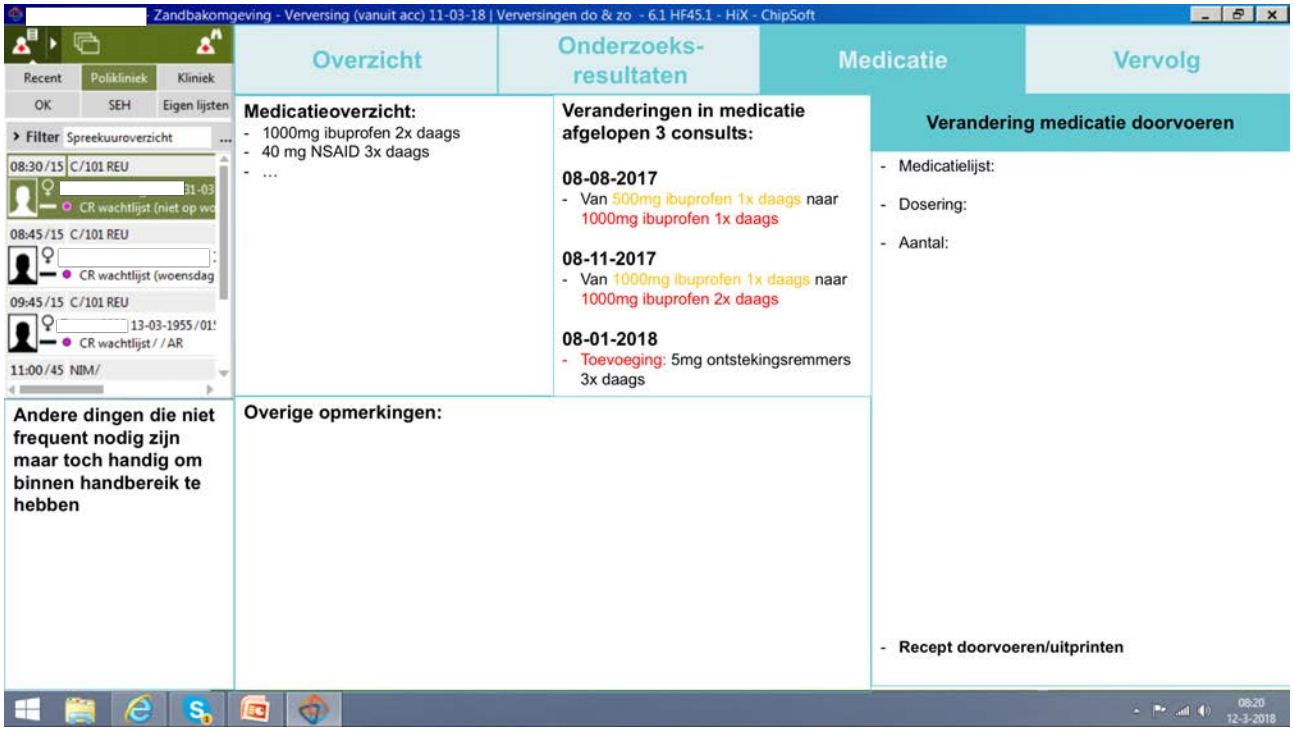

Figure 4.3: Dashboard page 3

Figure 4.4: Dashboard page 4

As can be seen, the prototype consists of four tabs. The first tab is an overview screen of the patient. As can be seen, on the left a list of patients is shown. It then displays the overview of the previous consult, as well as an overview of the current consult with a text box that includes all the standard questions asked. Below that a graph is placed which shows the disease course for the individual patient, as well as a chance on a flare according to the algorithm results. The right column is a column which includes all the data that needs to be measured in order to make the algorithm as accurate as possible. This includes the VAS score, the Swollen Joint count and the Tender Joint count, as well as the automated DAS calculator which is supposed to take the CRP values of the blood into account, along with the blood pressure values.

The second tab shows an overview of the results, both lab results and radiology results.

The third tab shows an overview of the medication, as to which medication the patient takes and the recent changes in medication that took place. Along with this a form is presented in the right column that the specialist can fill in to implement a change in medication. At the bottom a tab for additional comments is put.

The fourth tab concerns the continuation of the treatment, including the research that needs to be undertaken and a new appointment that needs to be made.

4.5.3.1 Choices made regarding design

Almost all of the requested functionalities as discussed in Chapter D.4 and Chapter 4.5.1 were implemented in this dashboard. This includes the automatic DAS calculator, better VAS representation, a nicer visual of the disease course, better overview of usage of medication, information required to be there at first glance, less clicking and a less cluttered dashboard, less manual things to be entered such as the list of questions. What was not embedded yet in the dashboard design was a pop-up about the eligibility for the research, since not enough information surrounding this topic was acquired. The design has been made in such a way that it guides the specialist in chronological order through all the necessary steps that are taken during an RA consult, as described in Chapter 4.5.2. From a usability perspective, it is very important that a dashboard guides and assists the user through all the steps taken [Chapter 4.4].

Since the most common inefficiency in the current dashboard is its complexity and its sheer number of tabs, a choice has been made regarding the design of the new dashboard to keep it simple and only stick to four tabs, directly placed at the top of the page where the users attention goes first, as mentioned in Chapter 4.4. A simple design has been proven to reduce cognitive overload [58].

Furthermore, the use of color has also been taken into account while designing the dashboard, since the color is more enhanced (a less transparent shade of blue compared to the rest) to draw the user's attention to that item, making it clear on which tab he or she is. In addition to this, a couple of visuals have been added to the dashboard. Fitting use of visuals in a dashboard has proven to increase clarity and overview, reducing cognitive overload and increasing efficiency (as seen in Chapter 4.4). These visuals includes the graph of the disease course, with underneath it a risk on a flare within a certain time span, where a chance of 0-40% would be visualized in green, 40-70% in orange and 70-100% in red. Same is done with the time span in which a flare can occur, within a week or two would be colored in red, three weeks up to 2 months in orange, and from 2 months upward in green. These colors show the risk like the traditional traffic light. More on this can be found in Chapter 4.4. This principle is also applied to the graph, where the bar next to the graph shows the intensity of the DAS value, and with that the flare, as well as to the puppets with swollen and painful joints, where painful joints is most of the times more bothersome to the patients. Another added visual is the VAS slider. As can be seen in Chapter 4.3, in the old dashboard only a number had to be inserted. However this did not give both the specialist and the patient enough clarity, as mentioned in Chapter D.4. The way the VAS is presented now clarifies the use and makes the question easier to answer, as well as the fact that it stimulates a patient-oriented treatment.

Grouping is also an essential ingredient (see Chapter 4.4). As can be seen in the above figures, the four tabs have grouped all the activities belonging to the four different parts of the consult. Attention has also been paid to this concept within the tabs. For example, at the overview tab, the entire right column includes all the tasks performed during the physical examination. Both overviews of the previous and the current consult are grouped together, and the flare information is all together in a box in the bottom.

4.5.4 Evaluation of first prototype

After the above prototype was developed, a meeting was scheduled with the stakeholder to evaluate this dashboard. A full transcript of our conversation can be found in Appendix D.6, including the answers to my questions. A lot of insights into the adjustments that needed to be made to improve the dashboard were generated. The first prototype was very well received, as it created more overview. Adjustments to be made on the first tab is that more room for the current consult was required, as well as some changes in terminology. Furthermore the painful and swollen joint puppets needed to be reversed. On the second tab, only some terminology needed to be changed. On the third tab, more room for the medication overview was required. Also a separate button for changing the medication would be preferable. On the fourth tab, some terminology needed to be changed along with a summary of the policy needed to be added, as well as an overview of the next appointments. Also, the suitability for research needs to be stated here as well as the ability to perform tasks like making an appointment and requesting lab research needs to be available. The full list of adjustments can be found in Appendix D.7.

4.5.5 Second Prototype

Based on the above accumulated opportunities for improvement, a second prototype was developed, which has these adjustment embedded in it. This prototype looks as follows:

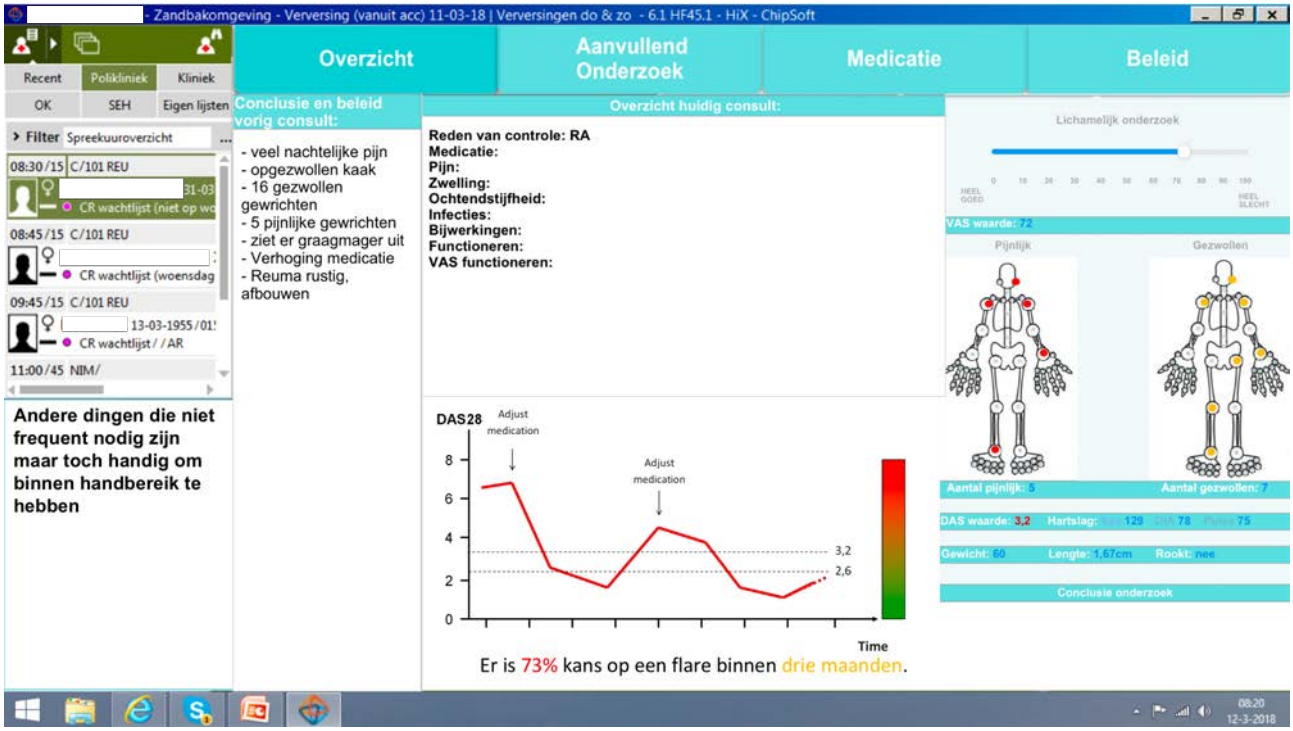

Figure 4.5: Dashboard page 1

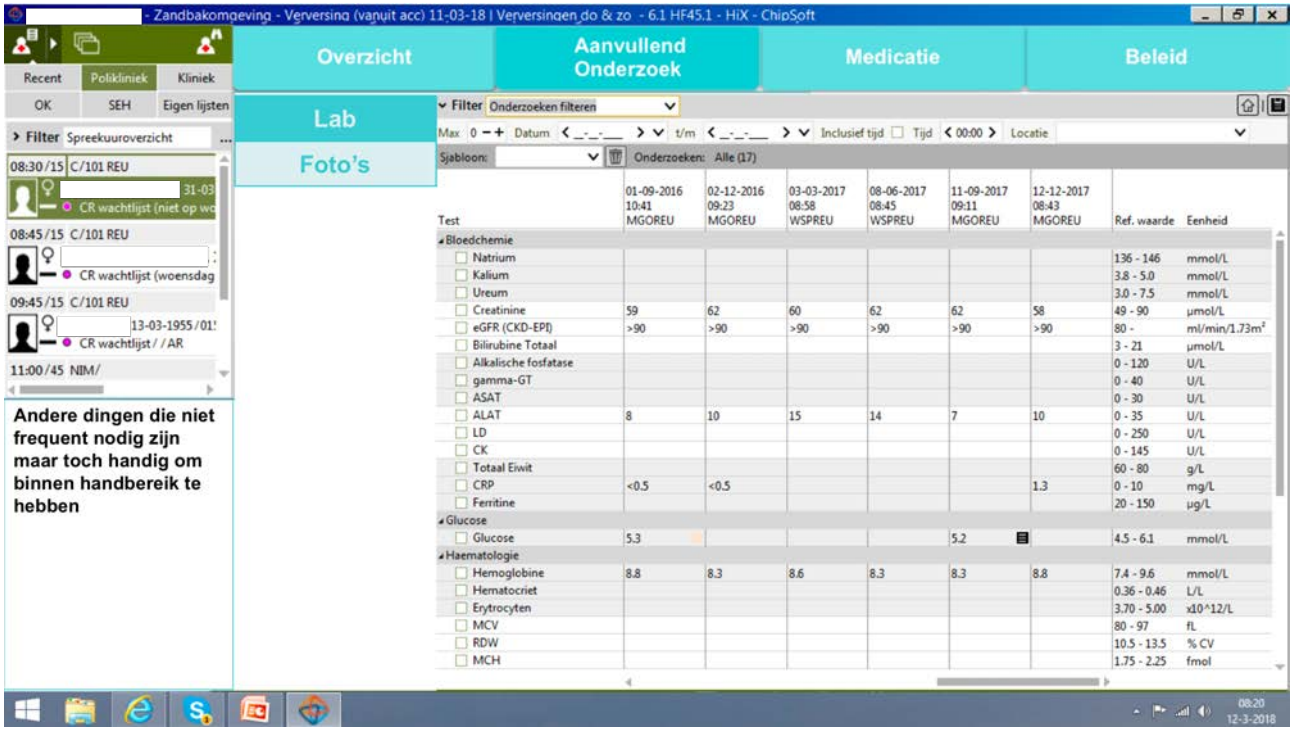

Figure 4.6: Dashboard page 2

Figure 4.7: Dashboard page 3

Figure 4.8: Dashboard page 4

4.5.5.1 Changes made compared to first prototype

The adjustments made to the four tabs are as follows. In the first tab, the overview of the current consult has been made larger, the right column has been made into a whole column that stands out compared to the rest, making it clear that this concerns the bodily research. Furthermore, the swollen and painful joint puppets have been swapped, putting the painful one first, since this results in the most unbearable effects, and the weight, length and smoking has been added. In addition to that, a text field about the research has been put at the bottom. The second tab has roughly remained the same, apart from the term radiology that has been changed to photos. In tab number three the medication overview has gained more space as per request. The change of medication has been changed in a button at the bottom, where, when clicked, a form will appear that will embed this change. In the final tab, the term "vervolg" or "continuation" has been changed to "beleid", or "policy". The arrangement of this tab has changed a bit, where a box with policy has been put where manually the proceedings of the treatment will be entered, a box that can also be filled before the consult by the specialist, depending on personal preference. Beside that, an overview of the next appointments is placed, so the specialist has an overview of what is planned and what needs to be planned. At the bottom of that box next appointments and a request for a lab research can be made by clicking the buttons. Furthermore, at the bottom of the page, it will automatically say whether a patient is eligible for further study, and by clicking on the "researchverpleegkundige" link, this will automatically be forwarded to the research nurse.

4.6 Future steps

A third prototype can be developed with sole focus on usability and design, not on content.

The dashboard may possibly be tested by a couple of RA specialists, to see if the dashboard meets all their requirements and to see whether it is conform usability standards.

Furthermore, a list of functionalities the prototype should be constructed, presented to the UMC, and implemented.

Chapter 5

Machine Learning Methods

In this chapter the research methods regarding the development of the machine learning algorithm are discussed. First the definition of a suitable machine algorithm will be given, secondly the methods for the development of the machine algorithm will be given. Finally, the testing procedure for the comparison of the joint modelling algorithm with the machine learning algorithm will be discussed.

5.1 Requirements of machine learning algorithm

First of all, a machine learning algorithm can always be improved and expanded. Especially in the scikit learn package which we chose, there are plenty of options. Therefore, we decided to define the scope of the machine learning algorithm in advance. We want to know how do standard machine learning algorithms compare to the joint modelling approach.

For this, we developed a machine learning algorithm with the use of Scikit and Pandas, choosing an independent approach to what has been developed in-house by the data scientists of the UMC Utrecht, so that the methods can be compared and evaluated independently.

We decided to make the predictions per patient. We wanted to see whether, based on the historical data, the patient will have a flare in the future. The data is structured as can be seen below in Chapter 5.2.1, Figure 5.2, taking the different courses of the patient into account. Since the data is structured as such, we decided to arrange it as seen in the blue figure shown below this, with the aim to create more overview and to have concrete results of which of the patients will, based on their historical data, flare up in the future. First, we made the choice to make use of feature selection that will later be used to fit the model, since feature selection provides an effective way to make predictions by removing irrelevant and redundant data. This can reduce computation time, improve learning accuracy and facilitate a better understanding for the learning model or data [72].

Afterwards, we chose the right machine learning method for the corresponding data, as elaborated on in

Chapter 5.2.3. We will test the prediction capabilities of the machine learning algorithm based on whether the model can make a prediction based on historical data for one date in the future. These methods will be further elaborated on in Chapter 5.3.

5.2 Development machine learning algorithm

In the following chapter, the steps taken for the development of the machine learning algorithm will be explained. First, the method as to how the data is prepared in order to get a good machine learning algorithm is discussed. Following this, features that might be useful in the prediction of the target, based on relevant research, are selected. These features will then be programmed and their influence on the prediction of the target variable will be determined. Finally, the machine learning algorithm will be developed by making use of several approaches.

5.2.1 Prediction of the model making use of data splits

As will be discussed in Chapter 5.2.3, three machine learning techniques have been used to predict whether a flare will occur with a patient. The goal of developing this machine learning algorithm is to be able to predict whether or not a flare will occur for a particular patient. To be able to do this, we have decided to split the data on a particular date, training the model in such a way that it can predict whether or not a flare will occur for a particular patient after this date. This has been done so that it can be seen how the model will perform when given a fixed time to make the prediction on.

The split date that has been chosen is 31st of December 2015, which we thought would be a good split since this divides the data set roughly in a 70 to 30% ratio, as can be seen in Figure 5.1.

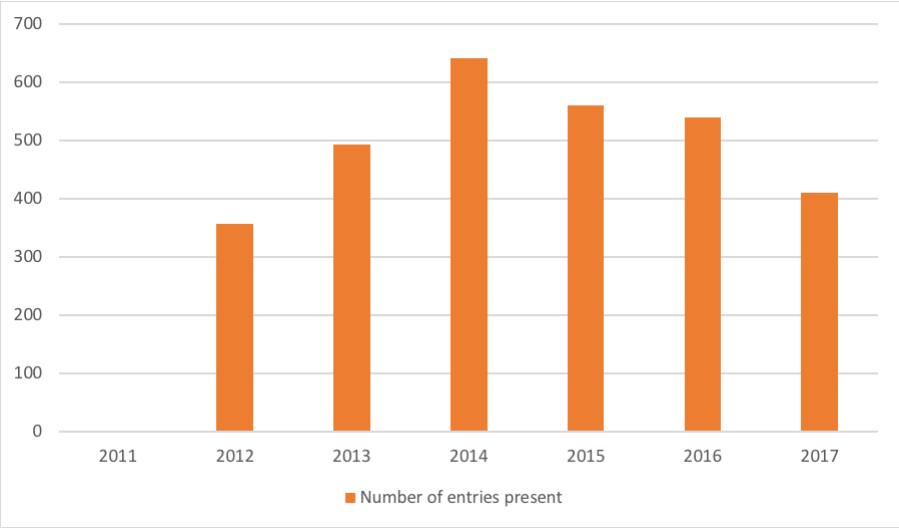

Figure 5.1: Number of registered events present per year.

Figure 5.2 shows the development process of the machine learning algorithm. First, the data set will be split into two different data sets, namely the data set before the split date, which includes all the records of the set that happened before and on the 31st of December 2015, and a data set after the split date that includes all the records after the 31st of December 2015. Second, features also called predictor variables will be constructed on the data set including all the records of before the 31st of December 2015, grouped per PseudoID, which is the unique ID of the patient. The data set of after the 31st of December 2015 will be prepared in such a way that it shows the PseudoID together with the matching target variable of that patient after the splitting date. Thirdly, the data set with the predictor variables before the split will be used to fit the three different machine learning techniques. This will be done making use of 10 fold cross-validation. These three machine learning techniques will afterwards generate a prediction for the target variable, predicting whether or not each patient will get a flare after the 31st of December 2015 or not. These results will be compared with the actual results of that period with use of the accuracy and the generation of a confusion matrix.

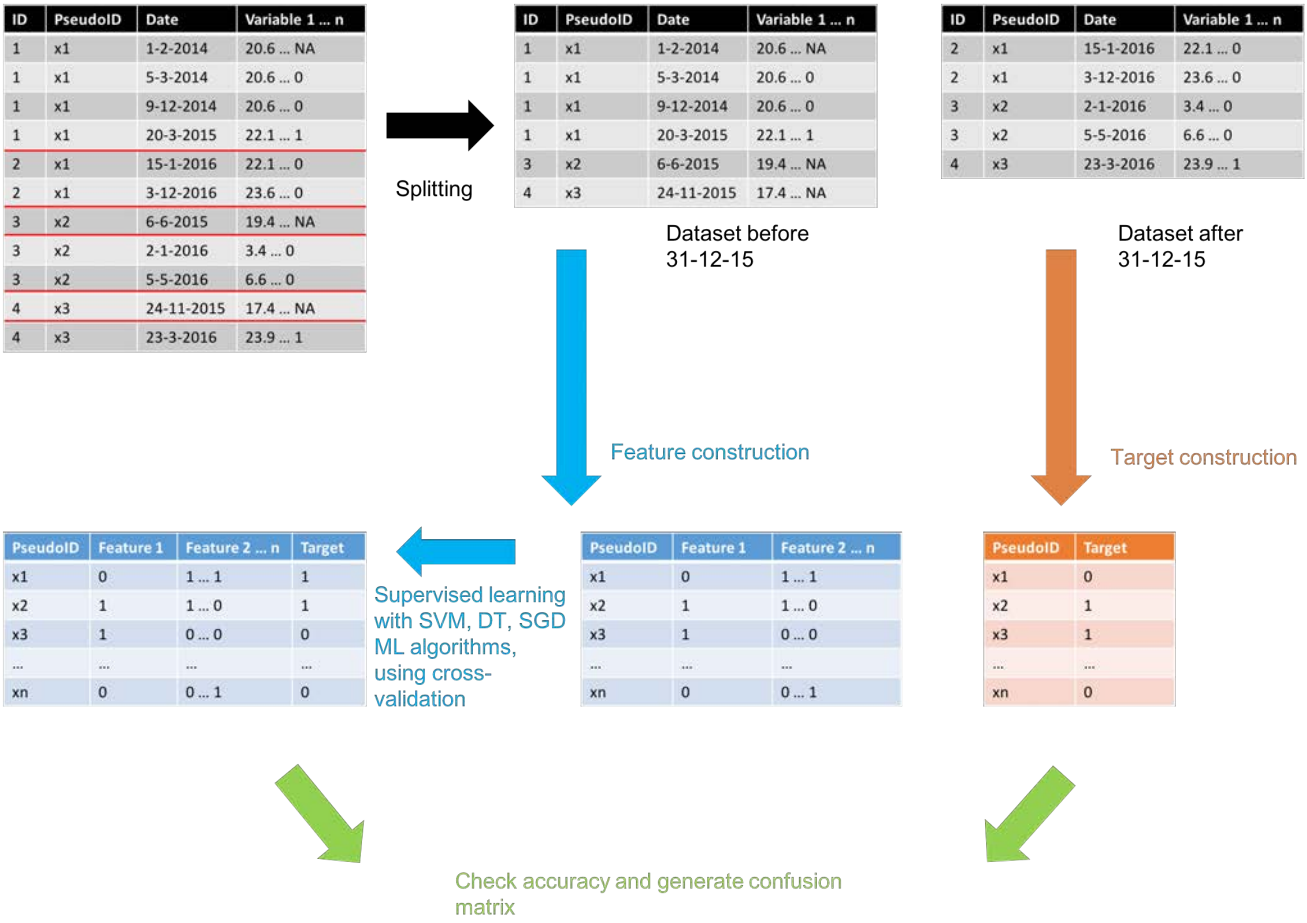

Figure 5.2: Development process of machine learning algorithm

For this machine learning algorithm, a choice has been made to predict the results for two targets. One is the target set by the hospital meaning whether or not a flare has occurred within a person who has just started

tapering. The other one is a wider target, that takes whether a patient will get a flare or not into account, regardless of whether a patient is tapering.

The data set before the split includes 283 unique patients, and the data set after the split includes 253 unique patients. However, the amount of unique patients that is present before and after the split date is 233. Therefore the confusion matrix will in total contain 233 results.

The above used method uses 10-fold cross validation. This is done to ensure that the actual reliability of the prediction is satisfactory and will be proportionate to predictions on other datasets. K-fold cross-validation is a technique that can be used to reduce the bias of overfitting on the current dataset, since it divides a data set in k equal subsets, uses each subset as a test set with k - 1 training sets and will be repeated k times and leads to k results. Two advantages of this method is that all data is used both as training data and test data and k tests of the desired performance indicator can be given instead of just one. Because of this, this method will give more insight into how reliable the model is [73].

5.2.2 Feature Construction

As a preparation for the development of the machine learning algorithm, a selection of features, also called predictor variables, has been made that have a high possibility to have an impact on the target variable. Based on the research done in Chapter 2.2.4 including a variety of influential factors on the occurrence of a flare, the following features have been constructed on the dataset before the 31st of December 2015. The full code of these features can be found in Appendix C.1.

1. **ChangeInMedication**: whether a patient has made a switch in medication;
2. **RiskGroup**: being in the RA risk group, meaning whether the patient has CCP and/or RF in the blood;
3. **Smoke**: whether a patient is a smoker or has smoked in the past;
4. **Taper**: patient started tapering;
5. **TJIncrease**: increase in number of tender joints;
6. **DoseDecrease**: there has been a decrease in dosage for a patient;
7. **DoseIncrease**: there has been an increase in dosage for a patient;
8. **BigJoint**: whether the patient has swollen big joints;
9. **SJIncrease**: increase in number of swollen joints.

Unfortunately, not all the influential factors such as weight, weather and stress could be taken into account while developing the features. This is due to a variety of reasons. Stress is not a measurable factor and weight, as described in Chapter 3.3, has only been recorded once. Weather is a relevant variable as shown in Appendix B.5 and could be taken into account for the time component of the prediction.

Several of these features were already constructed by the data scientists working on this project, namely the features tapering, DoseIncrease, Big Joint and SJIncrease.

The feature DoseIncrease puts a 1 if the dosage has increased for the patient and a 0 if this has not happened. The feature SJIncrease measures whether there has been an increase in the number of swollen joints, taking on the value 1 if this is the case. The feature tapering measures whether a patient has started a monitored reduction in the use of biologicals, turning 1 if this is the case.

All of the features that are created are of the boolean type, and are all applicable to the training set of the data.

The remaining six features were constructed from scratch, using the following methods: the **change in medication** feature looks at whether the number of unique ATC codes is higher than one over the entire course of the unique patient, up until the cutoff value of the training set. If so, the value will become one.

The **risk group** feature, also known as the feature where the patient has either CCP or RF in the blood, or both. This feature looks at whether either one of these elements occurs in the blood of the patient at any time. If so, the feature will turn 1, if this is never the case, the feature will be zero for the given patient.

The feature **smoker** checks whether a given patient is a smoker, or has been a smoker in the past. This was only occasionally recorded, as mentioned in Chapter 3.4, however if more of this data will be collected in the future, this could have a significant impact on predicting flares, as discussed in Chapter 2.2.4.

The feature that measures an **increase in tender joints** looks at whether or not the tender joint count has increased over time. If the tender joint count has increased over the recorded time, this feature will be one. If not, it will be zero.

The feature **DoseDecrease** checks whether the dosage has decreased over the given time for a particular patient. This feature is constructed based on another feature called Dosecat, which scales the dosage in dosage categories, with 1 being the lowest and 5 being the highest. If the difference in this category is lower than 0, then there has been a decrease in dosage and the feature will therefore take on the value 1.

The result of the constructed features can be seen below in Figure 5.3. The features TJIncrease and SJIncrease have the most significant impact on the target, followed by DoseIncrease and BigJoint. However, the correlation coefficient needs to be greater than or equal to 0.25 for there to be a correlation between the target and a feature [74]. The only features that conform to this rule are TJIncrease and SJIncrease, which both have a moderate to strong positive relationship with the target.

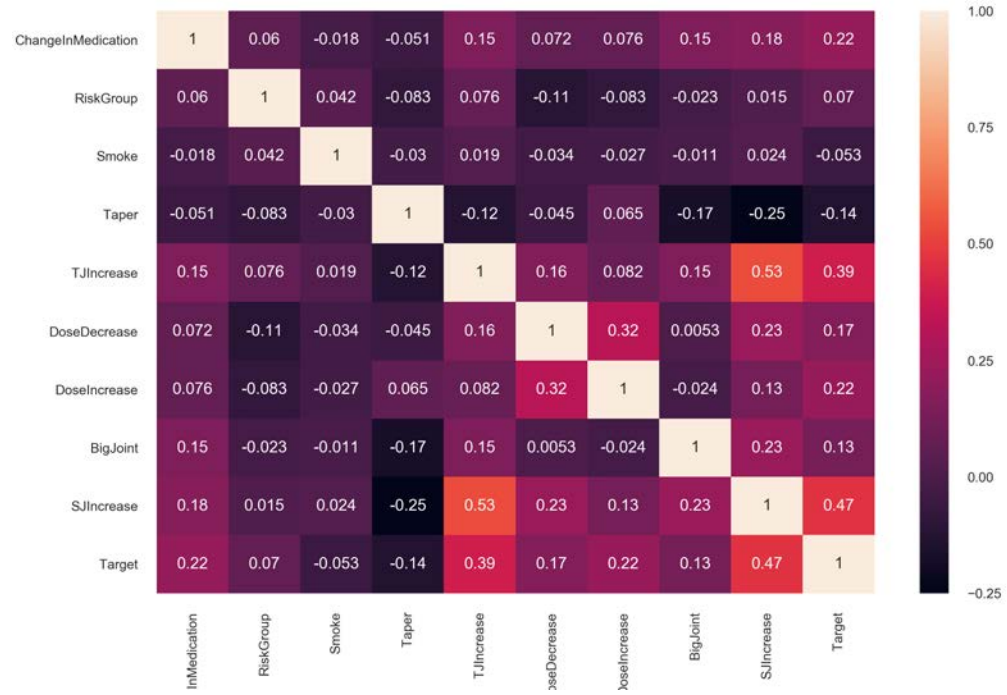

Figure 5.3: Correlation matrix of constructed features in relation to the target variable.

5.2.3 Machine learning approach

After the construction of the features, the right machine learning approach for this problem needs to be determined. The problem that we are facing is to be solved with supervised learning, since we are dealing with labelled data [73]. Since we are predicting a category, namely whether or not a flare will occur, classification is the right estimator to use [75].

To measure which classification approach would work best on this data, three classification approaches have been chosen to be fitted and compared, namely the Decision Tree Classifier (DT), the Support Vector Machine (SVM) and the Stochastic Gradient Descent classifier (SGD). In Chapter 6.1 the comparison of these of the techniques will be outlined.

The Decision Tree Classifier works as follows: it places the best attribute of the feature table at the root of the tree and repeatedly splits the training set into subsets. This process of splitting the training set into subsets is repeated until leaf nodes are found in all branches of the tree [77]. This tree is then used for the prediction done on the test set.

The Support Vector Classifier fits the data provided, returning a best fit hyperplane that divides the data. From there some features can be inserted into the classifier to see what the predicted class is.

Finally, the Stochastic Gradient Descent classifier is a method that implements regularized linear models with

stochastic gradient descent learning: the gradient of the loss is estimated each sample at a time and the model is updated along the way with a decreasing learning rate.

Each of the three above described methods has been fitted using the method as described in Chapter 5.2.1. Appendix C.2 contains the code for this method, for both targets as established in Chapter 5.2.1. Each of the methods is evaluated through the accuracy method, which is the proportion of data points correctly classified [76]. This accuracy is measured with the following formula: $(TP + TN)/n$, where n is the total number of instances in the correlation matrix.

5.3 Validation and Testing methods

Aside from the above comparison on accuracy of the three techniques, a confusion matrix of each of the methods will be generated and compared. A confusion matrix is a matrix that is used to measure the performance of a machine learning algorithm, specifically for supervised learning. This matrix consists of four cells, namely the true positives, false positives, true negatives and false negatives. The true positives is the number of data points correctly classified as having had a flare, the true negatives is the number of data points correctly classified as not having had a flare, the false positives is the number of data points incorrectly classified as having had a flare and the false negatives is the number of data points incorrectly classified as not having had a flare. As can be seen in Figure 5.3, a visualization of such a matrix is shown.

		Flare predicted	
		Y	**N**
Actual flare occurred	**Y**	True Positive	False Negative
	N	False Positive	True Negative

The above confusion matrix will be generated for each of the machine learning techniques, showing the performance of each. These results will then be compared with the confusion matrix generated by the joint modelling method.

5.3.1 Generation confusion matrix joint modelling technique

Since the joint modelling technique works differently than the machine learning method, a couple of adjustments needed to be made in order for the comparison between the two techniques to be valid. The joint modelling technique centers around episodes instead of patients. Each episode of has been cut off in the week of 31 December 2015 (the cutoff time), and that date has been used as the time-to-event in the training set. In the test set the actual time-to-event, which is the time that a flare occurred, has been set to the first week of 2016 onward. The point from which a prediction has been made have been matched with the actual visits and the period of prediction has been set to 12/24 weeks. These results have been grouped per PseudoID. With these settings, a confusion matrix has been generated.

Chapter 6

Results & Discussion

This chapter discusses the results of this research. Firstly, it discusses the results of the accuracy of the different machine learning techniques that have been used. After that, it discusses the performance of these different machine learning techniques. Thirdly, the joint modelling accuracy and performance results are discussed. Following that, the machine learning and joint modelling techniques are compared and finally the limitations of this research are discussed.

6.1 Results accuracy different machine learning techniques

In Table 6.1 the results surrounding the accuracy of the different machine learning techniques are displayed. The results include both targets chosen, the hospital and wider target, as described in Chapter 5.2.1. This accuracy is calculated using the accuracy score option, which computes the accuracy as the fraction of correct predictions, following the formula as described in Chapter 5.2.3.

	DT	SVM	SGD
Hospital target	60.1%	62.7%	70.0%
Wider target	54.1%	48.1%	57.9%

Table 6.1: Accuracy results of machine learning techniques.

As can be concluded from Table 6.1, the SGD classifier has the highest accuracy score for both of the targets, where all three models scored significantly lower on the wider target.

6.2 Results performance different machine learning techniques

As mentioned in Chapter 5.3, a number of confusion matrices have been generated to evaluate the performance of each of these machine learning techniques. Below these confusion matrices are visualized. The upper number in each box signifies the value for the hospital target, the lower value in each box represents the value for the wider target.

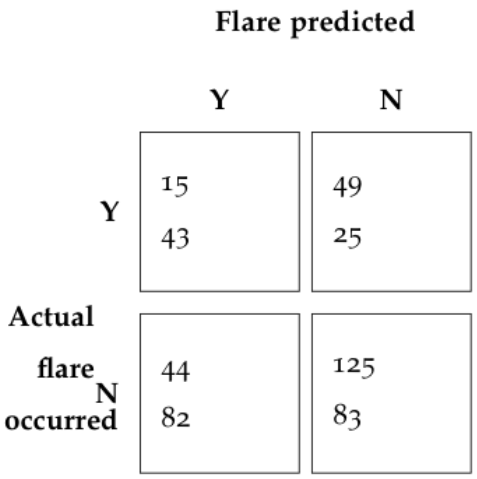

Figure 6.1: Confusion matrix decision tree.

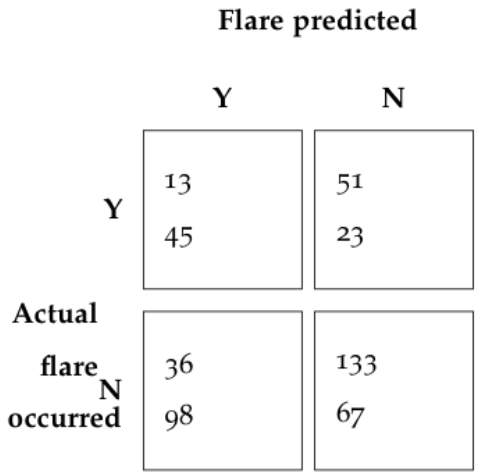

Figure 6.2: Confusion matrix support vector machine.

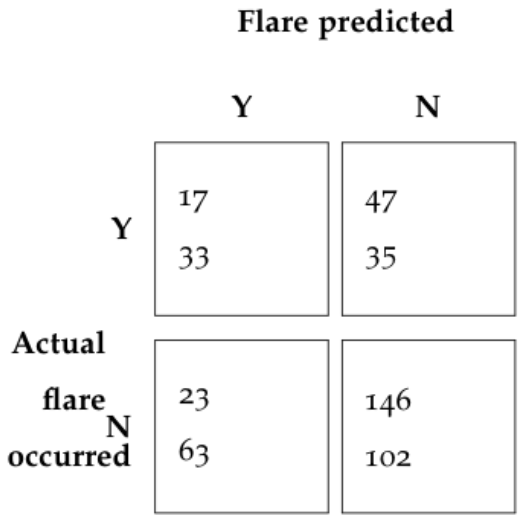

Figure 6.3: Confusion matrix stochastic gradient descent.

6.3 Accuracy and performance joint modelling technique

This chapter outlines the accuracy and performance results of the joint modelling technique generated from the methods as described in Chapter 5.3.1. These results are only generated for the hospital target, since this technique has only been developed for this target.

6.3.1 Accuracy

The AUC, which is the area under the ROC curve, for the joint modelling technique by only using the splitting data as criteria without using cross validation is 80.5%. The accuracy as calculated by the formula in Chapter 5.2.3, is 86%. The AUC for the joint modelling technique with the use of $5 * 3$ cross-validation, used without a fixed splitting date, is 77.4%. $5 * 3$ Cross-validation is a form of cross-validation that is single 5-fold cross-validation, used 3 times. In this method, the dataset that is used for the model fit, will be divided in 5 subsets. Of these 5 subsets, per fold 4 sets will be used to train the model and 1 set will be used to test the predictive performance. This is a technique as manually implemented in the joint modelling algorithm since there is an absence of possibilities within this approach for nested cross-validation.

6.3.2 Performance

Below in Figure 2.3 the confusion matrix of the joint modelling technique is displayed. The false negative rate for this method therefore is 0.369 and their false positive rate is 0.054.

		Flare predicted	
		Y	**N**
Actual flare occurred	**Y**	41	24
	N	10	175

Table 6.2: Confusion Matrix joint modelling

6.4 Comparison machine learning versus joint modelling

The performance of an algorithm or method is determined by how many instances are predicted correctly and incorrectly. However, it needs to be determined which of the measures are most important in this particular problem. A choice needs to be made as to what is worse: a false positive, so whether a flare is falsely predicted while it did not occur, or a false negative, whether a flare was not predicted but it did occur. Because of the severe uncomfort and pain a flare can give a patient, one can expect that it would be desirable that the model would produce as little as possible false negatives, especially in comparison the number of false positives. This decision is evidently up to the specialists and doctors.

Therefore based on the above, it would be useful to look at the false negative rates of each of the techniques and see how the various models perform. The false negative rate is defined by the following formula: $\frac{FN}{FN+TP}$. These rates are set out for all machine learning techniques. The results of this is shown in Table 6.3.

	Hospital target	Wider target
DT	0.766	0.368
SVM	0.797	0.338
SGD	0.734	0.515

Table 6.3: False negative rate for each method.

Based on the above results, the SGD method has the lowest false negative rate in comparison to the rest of the methods for the hospital target. For the wider target, the DT method has the lowest false negative rate.

6.4.1 Restrictions

Factors to take into account with comparison are that the joint modelling approach and the machine learning approach used are vastly different, since the joint modelling approach has a dynamic time element, focusing on the course of the patient rather than the individual patient on its own. Every time a course ends, meaning the patient has flared, a new course is started. This technique does not take the individual patient and its entire history as a focus point, rather it focuses only on the history of the preceding course, comparing it to similar courses of other patients. This is contrary to the machine learning approach developed in this research, which is an approach that predicts whether an individual patient has flared beyond a point in time, based on the entire history of the patient, taking the patient as a whole into account.

As mentioned above, the joint modelling technique has a dynamic time element, which causes the technique

to solely predict whether there will be a flare within a period in which the next consult will take place, which is 12-24 weeks. Therefore, the model had to be a small bit adjusted as described in Chapter 5.3.1 in order for the comparison to be valid.

Possible restrictions in the comparison are therefore the difference in methods, since the joint modelling method naturally does not predict using our approach, as well as the fact that the joint modelling method does not make predictions based on PseudoID, but rather on different courses. Therefore the model had to be adjusted to match the machine learning technique, taking the entire span of the patient into account.

Furthermore, with the machine learning method cross-validation is done on the model that uses data before the split. With the joint modelling method that takes the specific split into account this is not done. This could result in a reduction of the trustworthiness of those results and possible overfitting [73].

6.4.2 Comparison of results

Based on the above generated results and the results made available about the joint modelling technique, the following conclusions can be made. The joint modelling technique has a higher accuracy than our most accurate machine learning technique, with a difference between the two of 16% with the use of the splitting criteria and a difference of 7.4% with the use of the cross-validation by the joint modelling technique. However a possible factor that could have affected this as discussed in Chapter 6.4.1 is the difference in which the models were fitted.

The false negative rate of the machine learning methods on the hospital target were quite high in comparison to the joint modelling technique, showing that the joint modelling technique performs better in this regard. Both methods however show a significant difference in false positive rate compared to false negative rate, with the false negative rate being higher for both methods than their false positive rate. This is something that is quite unfavorable, as mentioned at the start of this Chapter 6.4.

An advantage of the joint modelling technique is that this technique is specifically designed for problems as encountered in this research, with the different metrics such as time-to-event. Yet, machine learning methods provide more insightful models, such as a decision tree, making them far easier to interpret than the joint modelling model. Since they provide more insightful models, possible errors in development can be easier detected. Machine learning methods also benefit from many functionalities that are embedded within the tools used to develop the model, like for example cross-validation, creating more overview in the code. This is not standard implemented in joint modelling, making the development of the joint modelling method possibly more time consuming.

6.5 Limitations

During this research a number of limitations were encountered. These limitations include the data quality, the unavailability of the machine learning algorithm as developed for the UMC Utrecht. If more of the necessary data would have been recorded, this could have improved the scores of the machine learning algorithm developed. Furthermore, the absence of the machine learning algorithm as developed by the external party in this research caused for the sub-question regarding the data science aspect to be answered not as fully as could have been possible when this was available. Since the machine learning algorithm developed for this research was subject to the constraints caused by the scope and time available for this research, it did not perform as well as supposedly a machine learning algorithm developed for at least a year would. This would have made the comparison between the two techniques more valid.

Chapter 7

Conclusion & Outlook

This chapter describes the conclusions, limitations and future work of this thesis. First, an overview of the research done will be given, along with an answer to the research questions. After that, the future work will be discussed.

7.1 Conclusion

The research conducted in this thesis consists of two main components namely the data driven prediction and visualization, which on the one hand is a data science problem, and on the other hand a usability aspect. Furthermore this research has an element that focuses on how these findings can be expanded to similar diseases. This will be discussed in Chapter 7.1.1.

The usability aspect of this research is *How to develop a dashboard for rheumatism specialists that improves their efficiency, contributes to better data collection and encompasses the results of the models generated?* The answer to this question was found by doing research into the usability field, focused on how to create a dashboard that promotes operational efficiency and improves usability. Furthermore, a number of visits with relevant stakeholders helped to identify the bottlenecks that reduce efficiency and usability in the current dashboard. With the use of the above components, a prototype has been made that contains all relevant criteria.

The data science aspect of this research has been outlined by the following sub-question: *How do standard machine learning algorithms compare to the joint modelling approach?* The development of a machine learning algorithm alone gave a lot of insights into how both techniques differ vastly. Even though the results show that the joint modelling method performs better, it has been shown that with the used methods a machine learning algorithm can be developed with a trustworthy accuracy and performance. Furthermore the machine learning method resulted in a clear and well interpretable model, namely a decision tree. It however remains difficult to determine which technique performs best given such situations, since a lot remains unknown surrounding the joint modelling techniques used. Both techniques should ideally be tested on a new unseen set of data.

7.1.1 Future work

The third sub-question of the research was focused around how this solution can be extended to other similar problems within the hospital. Both the data science aspect as the usability aspect of this research can be used to answer this.

As described in this research, both joint modelling and machine learning methods proved to be effective in the prediction of flares within rheumatoid arthritis patients. Similar methods could be used to expand to diseases with a similar course, such as inflammatory bowel disease, diabetes, lupus, multiple sclerosis and so forth. This can be done with use of the following ingredients: metrics need to be defined for these diseases, if not done already, that define based on various variables, like the DAS (see Chapter 2.2.3), when a flare occurs. In addition to that, enough time-series data needs to be recorded and available. Therefore, diseases with a similar course, like most autoimmune diseases, could qualify for expansion if metrics are present that define whether a flare occurs and if there is enough time-series data present to make solid predictions on. For this to work a system needs to be set up with regard to the collection of data, if not done already, similar to how it has been done with rheumatoid arthritis, that collects relevant data over a large period of time.

The success of the above is vastly tied to the usability aspect as mentioned in this research. The right collection of the data starts with the doctors, who will have to do this continuously in a uniform way. A tool that can significantly aid this process is the use of an intuitive dashboard, since it improves clarity as to where and how to collect the data.

Bibliography

[1] A. Monnappa, Data Science vs. Big Data vs. Data Analytics, May 2018. https://www.simplilearn.com/data-science-vs-big-data-vs-data-analytics-article. Accessed May 1, 2018.

[2] Difference between Data and Information, http://www.differencebetween.info/difference-between-data-and-information. Accessed July 11, 2018.

[3] F. Glynn, How to Personalize the Customer Experience With Big Data. https://www.allbusiness.com/personalize-customer-experience-big-data-101621-1 Accessed May 1, 2018.

[4] P. Nanetti et al, Economist Intelligence Unit, The Deciding Factor: Big Data & Decision Making. Capgemini whitepaper. https://www.capgemini.com/wp-content/uploads/2017/07/The_Deciding_Factor__Big_Data___Decision_Making.pdf. Accessed May 1, 2018.

[5] 3 advantages of using analytics. https://www.sfgnetwork.com/blog/business-intelligence/3-advantages-of-using-analytics/. Accessed May 1, 2018.

[6] B. Kayyali, D. Knott, S. van Kuiken. The big-data revolution in US health care: Accelerating value and innovation. https://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/the-big-data-revolution-in-us-health-care Accessed May 1, 2018.

[7] Desner Advisory Services LLC, Dec 2017. Big Data Analytics Market study. https://globenewswire.com/news-release/2017/12/20/1267022/0/en/Dresner-Advisory-Services-Publishes-2017-Big-Data-Analytics-Market-Study.html. Accessed May 1, 2018.

[8] B. Marr, Big Data In Healthcare: Paris Hospitals Predict Admission Rates Using Machine Learning. https://www.forbes.com/sites/bernardmarr/2016/12/13/big-data-in-healthcare-paris-hospitals-predict-admission-rates-using-machine-learning/#5828452979a2 Accessed May 1, 2018.

[9] Drs. M. van Velthuizen, Personaliseren van de zorg, internal report, Utrecht, February 2018. Accessed May 1, 2018.

[10] Prof. Dr. W. van Solinge, Eindresultaten en Vooruitblik ADAM, internal reports, Utrecht, February 2018. Accessed May 1, 2018.

[11] Format indiening pilot Applied Data Analytics in Medicine (ADAM), internal report, Utrecht, March 2017. Accessed May 1, 2018.

[12] UMC & ZO, Magazine voor UMC Utrecht medewerkers, februari 2018 nr 1. Neonatology, Psychiatry, EDEN and RA articles. P.4-9. Accessed May 1, 2018.

[13] NIMH RAISE Questions and Answers about Psychosis. https://www.nimh.nih.gov/health/topics/schizophrenia/raise/raise-questions-and-answers.shtml. Accessed May 1, 2018.

[14] What is Cardiovascular Disease? http://www.heart.org/HEARTORG/Conditions/What-is-Cardiovascular-Disease_UCM_301852_Article.jsp#.WxPeGlOFPFw. Accessed May 1, 2018.

[15] Rheumatoid Arthritis support network, what is rheumatism? [Online.] https://www.rheumatoidarthritis.org/ra/rheumatism/

[16] Technical Design document RA pilot project ADAM, Decision List RA project ADAM, Data Dictionary RA project, internal reports, February 2018. Accessed May 1 2018.

[17] 7 Ways to Avoid a Rheumatoid Arthritis Flare. https://www.everydayhealth.com/arthritis/rheumatoid-arthritis/7-ways-avoid-rheumatoid-arthritis-flare/. Accessed May 1 2018.

[18] M.E. Ellis, Signs of Rheumatoid Arthritis Exacerbations and How to Prevent Them. https://www.healthline.com/health/rheumatoid-arthritis-exacerbations. Accessed May 1 2018.

[19] Influenza vaccination of patients with systemic lupus erythematosus (SLE) and rheumatoid arthritis (RA) LJUDMILA STOJANOVICH Bezhanijska kosa University Medical Center, Belgrade University, Serbia. Clinical & Developmental Immunology, June-December 2006; 13(2-4): 373-375. DOI: 10.1080/17402520600800820.

[20] Shafay Raheel, Cynthia S. Crowson, Eric L. Matteson and Elena Myasoedova, Rheumatology, Mayo Clinic, Rochester, MN, 2Health Sciences Research, Mayo Clinic, Rochester, MN Smoking Is Associated with Unfavorable Flare/ Remission Pattern in Patients with Rheumatoid Arthritis. http://acrabstracts.org/abstract/smoking-is-associated-with-unfavorable-flare-remission-pattern-in-patients-with-rheumatoid-arthriti Accessed May 1 2018.

[21] T. Worth, Coffee, Alcohol, and More That May Affect RA. http://www.health.com/health/gallery/0,,20444194,00.html#smoking-9. Accessed May 1 2018.

[22] How Fat Affects Arthritis Being overweight can make arthritis, gout, lupus, fibromyalgia and other joint diseases and conditions worse. https://www.arthritis.org/living-with-arthritis/comorbidities/obesity-arthritis/fat-and-arthritis.php. Accessed May 1 2018.

[23] Dimitrov, S., Hulteng, E. and Hong, S. (2017). Inflammation and exercise: Inhibition of monocytic intracellular TNF production by acute exercise via beta2 adrenergic activation. Brain, Behavior, and Immunity, 61, pp.60-68. DOI: 10.1016/j.bbi.2016.12.017.

[24] Zautra, A., Burleson, M., Matt, K., Roth, S. and et al (1994). Interpersonal stress, depression, and disease activity in rheumatoid arthritis and osteoarthritis patients. Health Psychology, 13(2), pp.139-148. Accessed May 1 2018.

[25] Harvard Health Publishing, Understanding the stress response. https://www.health.harvard.edu/staying-healthy/understanding-the-stress-response. Accessed May 1 2018.

[26] Patberg, W. R., Rasker, J. J. Weather effects in rheumatoid arthritis: from controversy to consensus. A review. The Journal of Rheumatology July 2004, 31 (7) 1327-1334. Accessed May 1 2018.

[27] Gorin, A., Smyth, J., Weisberg, J., Affleck, G., Tennen, H., Urrows, S. and Stone, A. (1999). Rheumatoid arthritis patients show weather sensitivity in daily life, but the relationship is not clinically significant. Pain, 81(1), pp.173-177. DOI: 10.1016/S0304-3959(99)00010-X.

[28] Guedj, D. and Weinberger, A. (1990). Effect of weather conditions on rheumatic patients. Annals of the Rheumatic Diseases, 49(3), pp.158-159. Accessed May 1 2018.

[29] E. Lunardo. Rheumatoid arthritis triggered by estrogen and progesterone hormone imbalance in women: Study. https://www.belmarrahealth.com/rheumatoid-arthritis-triggered-by-estrogen-and-progesterone-hormone-imbalance-in-women-study/. Accessed May 1 2018.

[30] M. MacGill Arthritis and rheumatism: What's the difference? https://www.medicalnewstoday.com/art-icles/7625.php. Accessed May 1 2018.

[31] M.A. Dunkin, More Than Just Joints: How Rheumatoid Arthritis Affects the Rest of Your Body. https://www.arthritis.org/about-arthritis/types/rheumatoid-arthritis/articles/rhemuatoid-arthritis-affects-body.php. Accessed May 1 2018.

[32] Let's Dig Into Everything about RA https://www.rheumatoidarthritis.org/ra/, https://www.rheumatoidarthritis.org/ra/facts-and-statistics/, https://www.rheumatoidarthritis.org/ra/diagnosis/anti-ccp. Accessed May 1 2018.

[33] T. Worth, 8 Signs and Symptoms of Rheumatoid Arthritis http://www.health.com/health/gallery/0,,20464354,00.html#morning-stiffness-0. Accessed May 1 2018.

[34] Healthline, Swelling in Rheumatoid Arthritis, https://www.healthline.com/health/rheumatoid-arthritis-swelling. Accessed May 1 2018.

[35] Michael H. Weisman, Executive Series Editor, Daniel J. Wallace. Oxford American Rheumatology Library, Rheumatoid Arthritis. Accessed May 1 2018.

[36] What is rheumatoid arthritis? https://www.arthritisresearchuk.org/arthritis-information/conditions/rheumatoid-arthritis/what-is-rheumatoid-arthritis.aspx. Accessed May 1 2018.

[37] Porter D, Gadsby K, Thompson P, White J, McClinton C, Oliver S, DAS28 and Rheumatoid Arthritis: The Need for Standardization. Musculoskeletal Care. 2011 9(4):222-7. doi: 10.1002/msc.218.

[38] Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 1988;31:315—24. Accessed May 1 2018.

[39] How Biologics Treat Rheumatoid Arthritis. https://www.webmd.com/rheumatoid-arthritis/how-biologics-work. Accessed May 1 2018.

[40] DNAlytics, DAS28-CRP Calculators, https://www.rheumakit.com/en/calculators/das28_crp. Accessed May 1 2018.

[41] Riazzoli, J., Nilsson, J., Teleman, A., Petersson, I., Rantapaa-Dahlqvist, S., Jacobsson, L. and van Vollenhoven, R. (2010). Patient-reported 28 swollen and tender joint counts accurately represent RA disease activity and can be used to assess therapy responses at the group level. Rheumatology, 49(11), pp.2098-2103. Accessed May 1 2018.

[42] J. Marcin, R. Donovan, K. Watson, C-Reactive Protein Test, https://www.healthline.com/health/c-reactive-protein#high-crp-levels. Accessed May 1 2018.

[43] Asar, O., Ritchie, J., Kalra, P. and Diggle, P. (2015). Joint modelling of repeated measurement and time-to-event data: an introductory tutorial. International Journal of Epidemiology, 44(1), pp.334 to 344. Accessed May 1 2018.

[44] Joint modeling of longitudinal and survival data, Yulia Marchenko, Executive Director of Statistics, StataCorp LP, 2016 Nordic and Baltic Stata Users Group meeting. Accessed May 1 2018.

[45] Ibrahim, J., Chu, H. and Chen, L. (2010). Basic Concepts and Methods for Joint Models of Longitudinal and Survival Data. Journal of Clinical Oncology, 28(16), pp.2796-2801. DOI: 10.1200/JCO.2009.25.0654.

[46] O. Asar, J. Ritchie, P. A Kalra, P. J. Diggle, Joint modelling of repeated measurement and time-to-event data: an introductory tutorial. International Journal of Epidemiology, Volume 44, Issue 1, 1 February 2015, Pages 334-344, January 2015. Accessed July 12 2018.

[47] Rizopoulos, D. (2015). Comments on Joint modeling of survival and longitudinal non-survival data: current methods and issues. Report of the DIA Bayesian Joint Modeling Working Group. Statistics in Medicine, 34(14), pp.2196-2197.

[48] Hickey, G., Philipson, P., Jorgensen, A. and Kolamunnage-Dona, R. (2018). A comparison of joint models for longitudinal and competing risks data, with application to an epilepsy drug randomized controlled trial. Journal of the Royal Statistical Society: Series A (Statistics in Society). Accessed May 1 2018.

[49] Sweeting, M. and Thompson, S. (2011). Joint modelling of longitudinal and time-to-event data with application to predicting abdominal aortic aneurysm growth and rupture. Biometrical Journal, 53(5), pp.750-763. Accessed May 1 2018.

[50] I. Guler, L. Calaza-Diaz, C. Faes, C. Cadarso-Suarez, E. Giraldez, and F. Gude, Joint Modelling for Longitudinal and Time-to-Event Data: Application to Liver Transplantation Data, Computational Science and Its Applications - ICCSA 2014 Lecture Notes in Computer Science, pp. 580-593, 2014. Accessed May 1 2018.

[51] Peer Review #1 of Performance of joint modelling of time-to-event data with time-dependent predictors: an assessment based on transition to psychosis data (v0.1) 2016. Accessed May 1 2018.

[52] Stuart J. Russell, Peter Norvig (2010) Artificial Intelligence: A Modern Approach, Third Edition, Prentice Hall ISBN 9780136042594. Accessed July 11 2018.

[53] C. Anderson, Creating a Data-Driven Organization, https://medium.com/@sifium/machine-learning-types-of-classification-9497bd4f2e14. Accessed July 11 2018.

[54] S. Patel, Chapter 2 : SVM (Support Vector Machine) Theory, https://medium.com/machine-learning-101/chapter-2-svm-support-vector-machine-theory-f0812effc72. Accessed July 11 2018.

[55] V. Patolla, How to make SGD Classifier perform as well as Logistic Regression using parfit, https://towardsdatascience.com/how-to-make-sgd-classifier-perform-as-well-as-logistic-regression-using-parfit-cc10bca2d3c4. Accessed July 11 2018.

[56] A. Faiola, P. Srinivas, J. Duke, Supporting Clinical Cognition: A Human-Centered Approach to a Novel ICU Information Visualization Dashboard. AMIA Annu Symp Proc. 2015 Nov 5;2015:560-9. eCollection 2015. Accessed May 1 2018.

[57] van Engen-Verheul, M., Peute, L., de Keizer, N., Peek, N. and Jaspers, M. (2016). Optimizing the user interface of a data entry module for an electronic patient record for cardiac rehabilitation: A mixed method usability approach. International Journal of Medical Informatics, 87, pp.15-26. Accessed May 1 2018.

[58] M. Piernik. 8 ways to reduce cognitive load. July 2017. https://uxplanet.org/8-ways-to-reduce-cognitive-load-part-1-cc2048d1b157. https://uxplanet.org/8-ways-to-reduce-cognitive-load-part-2-4b0f9d8ef5ad. Accessed May 1 2018.

[59] Nielsen Norman Group. (2018). F-Shaped Pattern For Reading Web Content (original eyetracking research). https://www.nngroup.com/articles/f-shaped-pattern-reading-web-content-discovered/. Accessed May 1 2018.

[60] Pantazi, S.V., Kushniruk, A., and Moehr, J.R. The usability axiom of medical information systems. Int. J. Med. Inform. 75, 12 (2006), 829-839. Accessed May 1 2018.

[61] Ash, J.S., Berg, M., and Coiera, E. Some unintended consequences of information technology in health care: The nature of patient care information system-related errors. J. Amer. Med. Informatics Assoc. 11 (2004), 104-112; Accessed May 1 2018.

[62] Koppel, R. et al. Role of computerized physician order entry systems in facilitating medication errors. JAMA 293 (2005), 1197-1203; Accessed May 1 2018.

[63] Lee, D., Moon, J., Kim, Y. and Yi, M. (2015). Antecedents and consequences of mobile phone usability: Linking simplicity and interactivity to satisfaction, trust, and brand loyalty. Information Management, 52(3), pp.295-304.

[64] J. Nielsen Designing Web Usability: The Practice of Simplicity (1st ed.), Peachpit Press, Berkeley, CA (1999). Accessed May 1 2018.

[65] Nielsen Norman Group. (2018). Top 10 Guidelines for Homepage Usability. Availableat:https://www.nngroup.com/articles/top-ten-guidelines-for-homepage-usability/. Accessed May 1 2018.

[66] M. Soegaard, Usability: a part of the user experience. https://www.interaction-design.org/literature/article/usability-a-part-of-the-user-experience. The Law of Similarity - Gestalt Principles (1). Interaction Design Foundation. June 2018. https://www.interaction-design.org/literature/article/the-law-of-similarity-gestalt-principles-1. Accessed June 16 2018.

[67] Spiliotopoulos, T., Papadopoulou, P., Martakos, D. and Kouroupetroglou, G. (2010). Integrating Usability Engineering For Designing The Web Experience. IGI Global. Accessed May 1 2018.

[68] Wu, F., Chen, C., Lee, Y. and Chen, R. (2010). Effects of color sample display and color sample grouping on screen layout usability for customized product color selection. Computers in Human Behavior, 26(1), pp.51-60. Accessed May 1 2018.

[69] B. Mulholland, The Traffic Lights Of UX: Staying Smart With Color, August 2016. https://usabilitygeek.com/traffic-lights-ux-smart-color/. Accessed May 1 2018.

[70] 10 guidelines for great Dashboard design. http://www.uxforthemasses.com/dashboard-design/. Accessed May 1 2018.

[71] K. Whitenton. Minimize Cognitive Load to Maximize Usability. Nielsen Norman Group. December 2013. https://www.nngroup.com/articles/minimize-cognitive-load/ Accessed 16 June 2018.

[72] S. Saskialaa, S. Appavu, S. Geethac, Multi Filtration Feature Selection (MFFS) to improve discriminatory ability in clinical data set, Applied Computing and Informatics Volume 12, Issue 2, July 2016, Pages 117-127. Accessed July 3 2018.

[73] W. van der Aalst, Process Mining, Data Science in Action. Second Edition. Springer.

[74] R. Taylor, Interpretation of the correlation coefficient: a basic review. JDMS 1:35-39, January/February 1990. Accessed July 3 2018.

[75] Scikit-learn, choose the right estimator. http://scikit-learn.org/stable/tutorial/machine_learning_map/index.html. Accessed July 4 2018.

[76] Rickards CA, Vyas N, Ryan KL, Ward KR, Andre D, Hurst GM, Barrera CR, Convertino VA. Are you bleeding? Validation of a machine-learning algorithm for determination of blood volume status: application to remote triage. J Appl Physiol (1985). 2014 Mar 1;116(5):486-94. doi: 10.1152/japplphysiol.00012.2013.

[77] R. Saxena, how a decision tree algorithm works, http://dataaspirant.com/2017/01/30/how-decision-tree-algorithm-works/. Accessed July 4 2018.