2019•2020

Faculteit Industriële ingenieurswetenschappen

master in de industriële wetenschappen: nucleaire technologie

Masterthesis

Reduction of pre-treatment QA for intensity modulated plans using a complexity threshold

PROMOTOR :

Prof. dr. Brigitte RENIERS

PROMOTOR :

MSc. Alexandra JANKELEVITCH

COPROMOTOR :

ing. Kenny GEENS

Matthias Hermans

Scriptie ingediend tot het behalen van de graad van master in de industriële wetenschappen: nucleaire technologie, afstudeerrichting nucleaire technieken / medisch nucleaire technieken

Gezamenlijke opleiding UHasselt en KU Leuven

2019 • 2020

Faculteit Industriële ingenieurswetenschappen

master in de industriële wetenschappen: nucleaire technologie

Masterthesis

Reduction of pre-treatment QA for intensity modulated plans using a complexity threshold

PROMOTOR :

Prof. dr. Brigitte RENIERS

PROMOTOR :

MSc. Alexandra JANKELEVITCH

COPROMOTOR :

ing. Kenny GEENS

Matthias Hermans

Scriptie ingediend tot het behalen van de graad van master in de industriële wetenschappen: nucleaire technologie, afstudeerrichting nucleaire technieken / medisch nucleaire technieken

UHASSELT

KU LEUVEN

Deze masterproef werd geschreven tijdens de COVID-19 crisis in 2020. Deze wereldwijde gezondheidscrisis heeft mogelijk een impact gehad op de opdracht, de onderzoekshandelingen en de onderzoeksresultaten.

Acknowledgements

This master's thesis is the conclusion of the master's degree in industrial sciences: nuclear technology with a focus on medical nuclear techniques, which is a joint education of the UHasselt and KULeuven. With my background as a medical imaging technologist, I was very pleased to be able to conduct research in the radiotherapy department of the Jessa hospital in Hasselt: the Limburgs Oncologisch Centrum (LOC). During my time at the LOC I was able to learn variable skills and I gained experience that will last me a lifetime and will help me in my further studies and career.

First of I would like to thank my external promotors MSc. Alexandra Jankelevitch and ing. Kenny Geens for their guidance throughout the realisation of this thesis even though it was cut short by the Covid-19 crisis. I would also like to thank them especially for giving me the opportunity to present at the Belgian Hospital Physicist Association where I was introduced to a professional field that I'm looking forward to work in later myself. I would also like to thank the other staff at the LOC for the hospitable reception and always being available to answer any questions I had. I would also like to thank Prof. dr. Brigitte Reniers for the guidance to perform this research this academic year.

Last but certainly not least I would like to thank my family and friends for their continued support during my various studies to help me get through all of them. Even when the last months I was not able to see them during the crisis, I knew I could always count on them whenever I needed.

Table of content

List of figures

List of tables

Abstract

For all patients treated with intensity modulation at the Limburgs Oncologisch Centrum a pre-treatment EPID-based QA verification is performed to verify the deliverability of the irradiation plan. This research tries to predict the outcome of this verification based on a complexity metric in order to use this to reduce the amount of verifications on the linac itself.

To accomplish this, a computational script was developed with the Eclipse scripting API to determine the complexity of a treatment plan. Both a standalone version for multiple patient analysis and an Eclipse plug-in were made. The edge metric, which measures the open leaf sides of an aperture was used to quantify the complexity.

Positive linear correlations were found between the amount of monitor units and the edge metric, but also between the edge metric and the relative gamma passing rates for VMAT arcs planned for the Varian linacs. A threshold of 0.19 mm^{-1} was devised for prostate irradiations on a Clinac™ accelerator based on the gamma passing rates with 3%/1mm criteria. This threshold correctly predicts 22.1% of the plans that fail the pre-treatment verification with a false positive rate (FPR) of 1.52%. A threshold of 0.18mm^{-1} would correctly predict 36.5% of the failing plans with a FPR of 7.58%. Reoptimizing these plans before the verification can reduce the workload. On the other hand, it was not possible to single out plans that would certainly pass the pre-treatment verification.

Abstract (Nederlands)

Voor alle patiënten die in het Limburgs Oncologisch Centrum met intensiteit gemoduleerde bestralingsplannen worden behandeld, wordt voorafgaand aan de behandeling een op EPID gebaseerde QA-verificatie uitgevoerd. Dit om te verifiëren dat het plan correct toegediend kan worden. Dit onderzoek probeert de uitkomst van deze verificatie te voorspellen aan de hand van een complexiteitsmetriek om op basis hiervan de werkbelasting te verminderen.

Een script werd ontwikkeld met de Eclipse scripting API om de complexiteit van een bestralingsplan te bepalen. Zowel een standalone versie voor meerdere patiëntanalyses als een Eclipse plug-in werden gemaakt. De complexiteit werd gekwantificeerd door een '*Edge Metric*' die de open zijkanten van een MLC-opening meet.

Er werden positieve lineaire correlaties gevonden tussen het aantal monitor eenheden en de edge metric, maar ook tussen de *edge metric* en de relatieve gamma-criteria voor VMAT-bogen die gepland zijn voor de Varian-linacs. Er werden drempelwaardes van 0,19 mm^{-1} en 0,18 mm^{-1} voorgesteld voor prostaatbestralingen op een Clinac™-versneller op basis van 3%/1mm gammacriteria. De 0,19 mm^{-1} drempel voorspelt correct 22,1% van de plannen die de controle voor behandeling niet doorstaan met 1,52% foutpositieven (FPR). De 0,18 mm^{-1} drempel voorspelt 36,5% van de plannen correct met een FPR van 7,58%. Het heroptimaliseren van deze plannen voor de verificatie kan de werklast verminderen. Het was niet mogelijk om plannen uit te sluiten die zeker de controle voor de behandeling zouden doorstaan.

1 Introduction

Radiotherapy is an often-used treatment option for cancer. It can destroy a tumour or make it shrink so it is better fit to operate on. Due to the destructive effect of radiation, it is important that the tumour is the tissue receiving most of the dose to prevent side effects in healthy tissue. Because the tumour has to be irradiated correctly, quality assurance (QA) plays a big role in radiotherapy. Machine quality assurance makes sure that all the equipment works correctly, and all software and equipment is correctly calibrated. Pre-treatment patient specific QA verifies if the irradiation of the patient will be done as planned. Lastly there is the in-vivo dosimetry to measure during treatment that the correct dose is given to the correct tissue in the patient.

This research will focus on pre-treatment patient specific quality assurance of the irradiation plans. The purpose is to determine whether an irradiation plan can be delivered correctly to a patient as planned in the treatment planning software. For that reason, the Limburgs Oncologisch Centrum (LOC) has implemented a pre-treatment QA verification procedure to limit the chances that mistakes (e.g. data transfer problems, dosimetric errors etc.) can happen. To start, all the irradiation plans are made by a dosimetrist or physicist and afterwards verified by a colleague. When a plan is ready it will be irradiated without patient on the treatment unit that will be used during the full radiotherapy treatment of the patient. This is generally done the day before the first fraction is administered. An electronic portal imaging device or a separate ionisation chamber array registers the verification measurement. The measured plan will be compared to the calculated plan from the treatment planning system.

If a plan passes this pre-treatment QA verification, the plan is approved for treatment. On the first day of treatment, a final verification is done with the patient on the treatment couch. This is the transit dosimetry test, where the electronic portal imaging measures the exit dose. If this is ok, the treatment can continue.

The pre-treatment QA of the treatment plans will be the focus of this Master's thesis. The whole verification process takes up a lot of time for the physicists. From experience the physicists know that most plans pass this pre-treatment verification and that the plans that do fail are the ones with very irregular MLC patterns and high amounts of administered monitor units. That is where this research comes in, the purpose is to try and predict the outcome of the pre-treatment QA test by using a complexity metric described by Younge et al. [1], [2]

To try and predict if a plan will pass or fail the pre-treatment QA, an application script will be developed that can be plugged in to the treatment planning system to calculate the complexity of a treatment plan. A retrospective study will be performed to compare calculated complexities of previously planned radiotherapy treatments to their gamma passing rates of the pre-treatment quality assurance verification. If possible, thresholds will be determined to set a limit for which plans do not need the pre-treatment verification anymore because the complexity is lower than the threshold. A high threshold could be set for plans which need to be replanned because they would fail the pre-treatment verification. Both could result in the reduction of the workload from pre-treatment QA.

To start, a literature study is performed to provide more background on some topics. After that, the used metric will be described, together with the method and materials used during this research. In chapter 4 the results will be displayed and in chapter 5 the results and implementation of this metric well be discussed together with possible further research topics on this subject.

2 Literature

Cancer is one of the most common diseases in this current age. Every year 18 million people are diagnosed with cancer and 9,5 million people die because of it according to the World Health Organisation's numbers of 2018. [3] These numbers are even increasing. When looking at specifics in Belgium, in 2012 there were sixty-five thousand new cancer cases of which thirty-five thousand needed radiotherapy (RT) treatment. For 2025, the total number of new cases is estimated at 78 thousand and the total amount of RT courses needed would be around 42 thousand. Those radiotherapy courses include both curative and palliative care. These numbers show the growing importance of radiotherapy and thus the knowledge about it. [4]

Radiotherapy uses a beam of ionising radiation onto the tumour volume in a patient. The radiation will transfer the energy onto the cells it passes, both the healthy cells and those of the tumour. The ionising radiation will affect the DNA of those tissues it deposits energy to by altering the structure of the DNA causing faults in the DNA-strings. These will result in 2 major processes, it may cause apoptosis, also called cell death, which leads to the tumour shrinking in size, or it prevents the miotic process of the cell so it can not proliferate resulting in a stagnant size of a tumour. [5]

As told in the previous paragraph, those effects will happen in both the tumour volume and the healthy tissue. That is why the healthy tissue needs to be spared as much as possible during radiation. An important factor in that process is the different response of tissues to a certain dose of radiation. The same dose to different types of tissues gives a different biological effect. [5]

2.1 Different treatment techniques

The required dose of radiation for the treatment can be delivered by means of two main techniques. Internal therapy where the radiation comes from within the body due to pills of an active material that patients swallow, injection of radioactive fluids or brachytherapy. This way of treatment gives a more local dose to the tumour with less damage to surrounding healthy tissue. Brachytherapy, where sources are permanently or temporarily implanted into the body, is an invasive procedure. It is also not possible to use this technique everywhere in the body, in the brain for example this can not be used.

The most used treatment method on the other hand is external beam radiation therapy (EBRT). In EBRT, the source of the radiation is outside of the body. This source is a linear accelerator (linac) where a beam of particles with high energy is generated. The beam of high energy particles delivers the radiation dose both in tumour volume and healthy tissue. The particles that are being used can consist of high energy photons, electrons, neutrons, protons or heavier ions, the latter two are often grouped under the name of hadron therapy.

In this paper, different treatment methods with intensity modulated dose will be examined, all using high intensity photons: intensity modulated radiotherapy (IMRT), volumetric modulated arc therapy (VMAT) Stereotactic radiosurgery (SRS) and stereotactic body radiation therapy (SBRT). These techniques differ from three-dimensional conformal radiotherapy (3D-CRT) in that they use different intensities in one bundle, while 3D-CRT has a fixed intensity for one irradiation beam.

2.1.1 Intensity modulated radiotherapy (IMRT)

IMRT uses multiple beams at different static angles to irradiate the tumour, just like conformal radiotherapy does. The difference is that IMRT uses intensity modulated beams, this means that a beam from one angle will irradiate the patients with intensities that can be different for every control point in the beam. This is caused by the movement of the leaves in the multileaf collimator (MLC). There are two techniques to acquire this modulated intensity: sliding window and step-and-shoot. The step-and-shoot method uses multiple discrete apertures formed by the MLC leaves at one angle. The linac only delivers the beam when the MLC leaves are not moving. Sliding window uses a continuous beam instead, where the MLC leaves move while the tumour is being irradiated. The movement of the MLC leaf pairs allows for better distribution of the dose in the body: more dose to the tumour while delivering a lower dose to healthy tissue and organs at risk (OAR). [6], [7]

2.1.2 Volumetric modulated arc therapy (VMAT)

While both conformal radiotherapy and IMRT are using static angles to deliver multiple (intensity modulated) beams to a tumour, the VMAT technique uses a beam that continuously rotates around the patient to form an arc. Multiple arcs can be used in one treatment plan for a patient. The multi leaf collimator continuously shapes/modulates the beam while the linac is moving along the arc like the sliding window technique of IMRT. The dose rate and gantry speed can also be changed during the treatment of the arcs. The purpose is to get the best conformal dose distribution and therapeutic ratio as with IMRT. The advantage of VMAT compared to IMRT is that VMAT generally needs a lower total amount of monitor units in one fraction to acquire the same dosimetric quality as IMRT. Lower amounts of MU result in less scatter and leakage. This results in lower doses absorbed in the tissue outside of the treatment volume, which reduces the risk of secondary malignancies. VMAT treatments will result in a dose that is more spread out over the tissue around the tumour, where IMRT still has parts of the body that receive a high dose because of the fixed treatment angles. [6], [8], [9]

2.1.3 Stereotactic body radiotherapy (SBRT) and stereotactic radiosurgery (SRS)

Both techniques are quite similar, compared to the previously discussed techniques it delivers a higher dose to a tumour during each fraction of a treatment. This high dose leads to the ablative tumour effect to achieve permanent disease control. VMAT and IMRT are mostly used, but fixed beams are also a possibility. SBRT and SRS differ from each other by the place in the body they treat; SRS is used to treat the elements of the central nervous system, so mostly tumours in the spine and skull, while SBRT is used for tumours elsewhere in the body. Both SRS and SBRT require very high precision of the dose delivery because of the high dose that is delivered in one fraction. Small margins are used to irradiate the healthy tissue as little as possible because the dose and dose gradient are high. A small deviation can result in a big difference when it comes to the treatment of the volume and possible side effects in healthy tissue. To achieve this precision SRS treatments in the brain use masks and markers to assign 3D-coordinates to the tumour and make sure that the head stays in the same place. SBRT relies more on image guiding and body frames to keep the body in the correct place. Depending on the position of the tumour in the body, gating is also applied, this monitors the respiration of a patient to determine when to irradiate the tumour. [10]–[12]

Most of these treatments are done in one single fraction, although sometimes multiple fractions are used with lower dose per fraction than single fraction SRS. This is done in bigger tumours to reduce the dose per fraction delivered to healthy tissue. Because the volume of healthy tissue receiving a high dose increases proportionally with the size of the tumor. Using multiple fractions allows the healthy tissue to heal in between fractions. The administered dose in these fractions is still high enough to achieve the ablation effect. Because SRS and SBRT use a high dose, a steep dose fall-off is needed outside the target to protect healthy tissue. This also results in less uniform target dose distributions compared to normal irradiation plans which use more fractions with lower amounts of MU’s and a more uniform dose

distribution. Heterogeneities with a high dose might however offer an advantage in the treatment of radioresistant cells. [10]–[12]

2.2 Treatment planning system

To put all the possible treatment techniques to use, a treatment planning system is needed (TPS). The calculations to find the best possible irradiation plan are not done by hand but with the treatment planning system software that calculates different ways to deliver the prescribed dose to the defined area inside a patient. The TPS tries to simulate a plan that delivers most of the dose to the tumour with a homogeneous distribution in the PTV while also sparing the organs at risk as much as possible to reduce the side effects that can occur due to radiotherapy.

The treatment planning system calculates an irradiation plan in two major ways: forward and inverse planning. Both methods are restricted by the physical capacities of the linac: the movement relative to the couch position, the movement limitations of the jaws, wedges and MLC leaves, both in position and speed. Forward planning requires the operator to put in most of the parameters manually, this means choosing the angles at which beams will irradiate the patient, the energy of the beam(s), the motion of the leaves in the MLC, jaws, wedges etc. After all the parameters have been put in the TPS, a dose distribution will be calculated. This is a trial and error process where more experienced workers will manage to plan the treatments faster.

That method is limited to the parameters that are put into the system. The TPS has little room to adjust the plan. That is what inverse planning does, instead of putting all the information about the linac in the TPS to see what dose it gives, operators put in the dose constraints for the tumour, organs at risk and other healthy tissue. The TPS will then calculate the parameters of the linac: the angles at which it has to irradiate, the MLC positions, beam intensities etc. It is difficult to get the perfect plan because that would be 100% dose to the tumour and 0% to the surroundings. That is why there are weighting factors for each dose constraint, so the software knows what constraints it certainly must accomplish and which it can let go of if it is not possible to comply with all the constraints.

The dose calculation is done based on the computed tomography (CT) images that are taken in the exact position a patient will be irradiated on the linac. These images come with a voxel map of Hounsfield units, this describes the radiodensity (whether radiation can easily get through the tissue or not) of a part in the human body. The higher the number the higher the electron density of the respective cells in the body. To correctly assign the densities to the images of the body, the CT must be calibrated. This is done with an oval phantom that represents a body and inside that phantom are multiple cylinders which represent different types of tissue found in the human body (bone, lung, muscle …) of which the density is known. This phantom is scanned on the planning CT scanner after which a calibration curve can be made between the electron density units and the Hounsfield units from the images of the phantom. [13]

2.3 Quality assurance procedure

In radiotherapy, quality assurance (QA) is of high importance to assure that the correct dose is delivered to the designated area in the body. Otherwise, the radiation therapy can not correctly affect the tissue that needs treatment, resulting in an uncomplete and ineffective irradiation. Instead, the healthy tissue could get more dose which can lead to negative side effects like diarrhoea, nerve damage and radiation-induced tumours.

There are multiple types of quality assurance, there is machine QA, which focusses on the quality of the machines and the software used to plan, deliver, and control the irradiations. This is an important part of the QA-process, it increases the chance of detecting errors and thus reducing the chance that errors will affect patients. Systematic errors of the machines or software could mask errors normally detected by the second type of QA: patient specific quality assurance. In patient specific QA the irradiation plan

of a single patient is tested to make sure that the planned treatment can be delivered to the patient correctly and that this continues throughout the complete treatment process. [14]

With the introduction of more complex treatment methods like IMRT and VMAT, the importance of QA has grown. There are more elements of the linac that can affect the outcome of the irradiation, like the movement of the MLC leaves during the irradiation (both the speed and position are important), gantry speed, dose rate… All these elements introduce possible errors and the more complex a plan is made, the more chance there is for these errors to happen. With the introduction of inverse planning the system has more freedom to make these complex plans with lots of different angles, speeds and intensities. [15]

Hence patient specific QA is important for these intensity modulated plans. Many parameters influence the outcome and due to the way of calculating the plan, this is different every time for every single patient. It is important that this verification is done before the patient starts with the first irradiation treatment, to rule out errors when the patient is on the table. Besides the pre-treatment QA-test there is also need for quality assurance in vivo, where the patient is on the couch receiving the planned irradiation and the detection of the beam is done by the EPID. Transit dosimetry tests are performed to make sure that the treatment plan is also correct with the patient on the table. [16], [17]

2.3.1 QA measurement devices

2.3.1.1 Low-resolution dosimetry devices

There are different ways that the pre-treatment, also called pre-delivery, QA tests are performed. The diode and ionisation chambers are well known devices to measure the dose. But they come with some downsides. The measurements with diodes are angle dependent, which means that the diodes should turn together with the gantry to get correct results and counteract the angular dependency. Another solution is to use biplanar diode arrays, they consist of two 2D diode arrays orthogonal from each other. Ion Chambers on the other hand have their finite size as a drawback. This results in averaging of the dose over the volume that is being examined seeing it as a homogeneous region, when in reality it is usually not homogeneous. By using ion chambers and diodes in detector arrays multiple points can be measured at the same time. The resolution of these arrays is not always sufficient to detect small MLC leaf errors, which are important in highly modulated IMRT and VMAT irradiation plans. [15], [18], [19]

2.3.1.2 High-resolution dosimetry devices

To get a better spatial resolution to detect those small MLC leaf displacements, film- and electronic portal imaging device (EPID) dosimetry are two possible techniques. Both have spatial resolutions of 1 mm or even better. When using films, every film must be calibrated, and read out by removing them from the setup used for verification, usually this is a volume of tissue equivalent material like Perspex in which films can be inserted. The use of films has some drawbacks, most types of radiochromic films are single use only, meaning that after every beam they have to be replaced which means that a lot of films are needed and have to be read out. Another downside is that they can not be used for real-time dose measurements, or to give a trend of the dose in time. Although there is research being done to make reusable films and develop techniques to read out films in real time. [20]–[22]

2.3.1.2.1 Electronic portal imaging device

Electronic portal imaging devices are part of the modern linacs. Their first use was solely for visualisation and control of the patient setup. These images are made by the megavoltage (MV) photon beam from the linac. The first generations of EPID’s ware based on liquid ionization chamber arrays and phosphor scintillating screens witch analogue or CCD cameras. The quality of those images was bad compared to the current standards and it was not implemented for dosimetric usage because of the limited dosimetric characteristics. This changed when the newer generation of EPID’s was introduced: the amorphous silicon electronic portal imaging device (a-Si EPID).

These a-Si-EPID's consist of multiple layers to produce a signal that can be used to visualize the position of the patient and give dosimetric information. The first layer is a metal/phosphor layer, the metal is used as build-up to increase the detector efficiency by helping in the conversion of photons to electrons to deposit the energy, this layer is more important for high energy beams (above 6 MV). The phosphor layer (Gd2O2S:Tb) converts the deposited radiation energy to optical photons in the visual electromagnetic spectrum. These optical photons are detected by the amorphous silicon layer that serves as a semi-conductor. They are converted into an electrical current. This current is what forms the two dimensional (2D) digital image of the patient, phantom or beam and can lead to dosimetric information after a specific calibration. The digital images can contain image lag and ghosting artefacts, these effects are strongest with low irradiation times, below those of routine clinical use.[16], [23]–[26]

2.3.2 QA evaluation

To evaluate the patient specific treatment plans a method is needed to compare the measured pre-treatment verification from one of the detectors to the calculated irradiation plan of the treatment planning system. Low, Harms, Mutic and Purdy of the Mallinckrodt Institute of Radiology introduced a quality measurement called the gamma index (γ) to determine this difference between the calculated and measured dose based on two acceptance tolerances. This is used to perform quantitative assessments of the quality of the calculations.[27], [28]

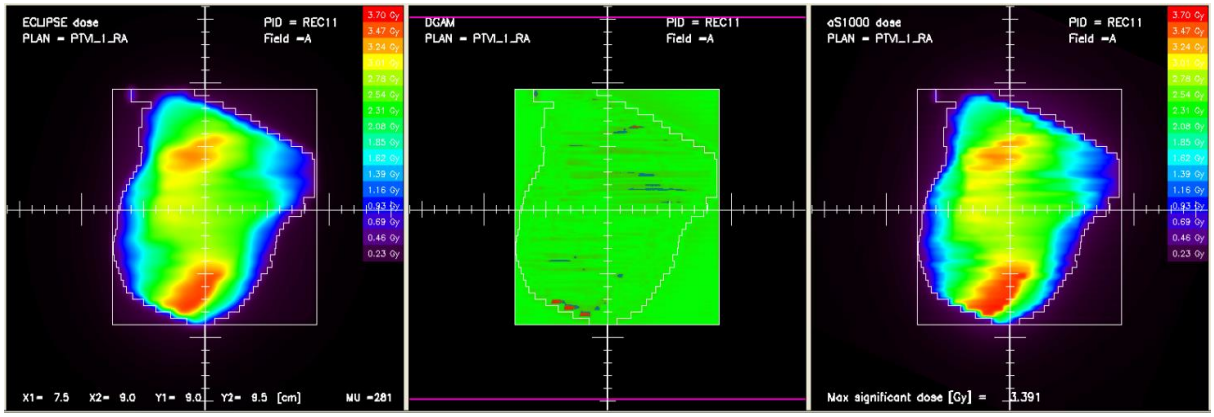

Figure 1 Example of a gamma analysis with the calculated dose map on the left, the measured dose map on the right and the gamma comparison in the middle. The gamma index for these dose maps is 98.95% for 3% DD and 3mm DTA. [28]

The first acceptance criterion is the dose difference (DD), this is used in regions with low dose gradients. Calculated and measured doses are compared directly, the tolerance is a chosen percentage that they may differ from each other. The second criterion is the distance to agreement (DTA), this is used in high dose gradient regions. The DTA tolerance is a distance in millimetres that represents the maximum distance the nearest point of a calculated dose can be away from the measured dose of the verification plan within the margin set by the DD. In high dose gradients, a small spatial displacement results in a large dose difference. Because they operate in different dose gradients, they complement each other to be used in one overall parameter. [27]

Figure 2 is a visual representation of the combined dose difference and distance to agreement parameters. The shaded circle represents the DTA parameter with Δd_M as the radius and the criterion of acceptance. The δ-axis represents the DD parameter with ΔD_M as the criterion of acceptance. The γ-index is calculated for every measurement point r_m relative to the surrounding calculated points r_c. The calculation of every single γ-index is done with the following formulas:

$$\Gamma(r_m, r_c) = \sqrt{\frac{r^2(r_m, r_c)}{\Delta d_M^2} + \frac{\delta^2(r_m, r_c)}{\Delta D_M^2}} \quad (1)$$

where

$$r(r_m, r_c) = |r_c - r_m| \quad (2)$$

and

$$\delta(r_m, r_c) = D_c(r_c) - D_m(r_m) \quad (3)$$

Formula 1 represents the surface of the ellipsoid formed by the DTA and DD acceptance criteria. The gamma index ($\gamma(r_m)$) of the measurement point is the minimum value of the calculated $\Gamma(r_m, r_c)$ for every calculated control point r_c. If this minimum value is below or equal to 1 the calculation passes the test, if it is above 1 it fails. This calculation is done for all the measurement points r_m and a percentage is given for how many points passed the gamma index test. [27], [29]

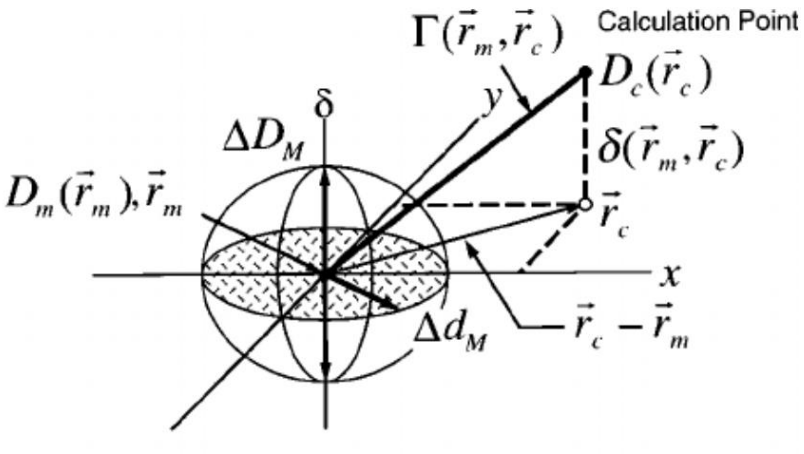

Figure 2 Geometric representation of dose distribution evaluation criteria using the combined ellipsoidal dose-difference and distance-to-agreement tests. [19]

For a plan to pass this gamma evaluation, the percentage of points that passed the test must be above a previously determined percentage. It is also worthwhile to analyse the dose differences visually, look at the average γ and the highest γ percentile. The actual constraints and minimum required pass rates can differ between different treatment methods. *[27], [29]*

Low et all suggests the use of 3%/3mm criteria when evaluating treatment plans with the gamma test. When the constraints are stricter, the required percentage to pass the test will be lower as well, this is for the both the local and global gamma thresholds. The global gamma threshold removes points in the out of field region which deviate more than that set percentage from the maximum dose in the dose distribution. This is done to prevent that gamma calculations in the out of field region skew the gamma test results. The actual constraints that have to be used depend on the used measurement device, because they have different accuracies. 2D or 3D analysis can also change the constraints, those of 3D analysis are often more strict because of the higher amount of measurement points that are analysed. *[15], [27], [29]–[31]*

2.4 Predictability of QA-evaluation.

With the introduction of the complex intensity modulated plans into the workflow and the growing use of them, the workload of the physics department has increased because of all the patient specific pre-treatment QA. That is why there is more and more research being done on the topic of predicting the outcome of these quality assurance verifications. Since most plans pass the QA testing, finding a method to identify plans which would pass the QA verification and which have a higher chance of failing it, can be beneficial for the workload. In a research conducted in most radiotherapy centres in the UK, 80% of the centres reported high modulation as the most common reason for out of tolerance measurements. [32] The TPS can make the irradiation plans overly complex, certainly with inverse planning, resulting in more discrepancies between the calculated dose and how it is irradiated and measured. Therefore, research on predicting the QA outcome focusses on finding complexity metrics (CM) that best represent this complexity related to failing QA tests. Several complexity metrics have been developed with varying results in predictability. The following paragraphs describe some examples of how this complexity can be calculated.

2.4.1 Modulation complexity Score

The modulation complexity score (MCS) uses multiple parameters from the treatment planning system to try and predict the deliverability. It is based on the variability in leaf positions, the degree of irregularity in field shape and the weight of each segment of a beam. This complexity score is divided into two parameters that get multiplied by the weight (the monitor units of a segment divided by the total amount of monitor units of the beam), this multiplication is done for each segment of a beam, afterwards the individual values of the segments are summed. [33]–[35]

$$\mathrm{MCS}_{beam} = \sum_{i=1}^{I} LSV_{segment\ i} \times AAV_{segment\ i} \times \frac{MU_{segment\ i}}{MU_{beam}} \tag{4}$$

The leaf sequence variability (LSV) is the first parameter that characterises the variability in the segment shape of a plan. The change of position in adjacent leaves is calculated and subtracted from the maximum position difference of the leaves in one bank for that segment. This is calculated for all the open leaves and summed up. The sum gets divided by the multiplication of the number of open leaves and the maximum position difference of the leaves. The calculation is done separately for each leaf bank of a segment and then the scores of both banks are multiplied by each other.[35]

$$LSV_{segment} = \left\langle \frac{\sum_{n=1}^{N}(pos_{max} - (pos_n - pos_{n-1}))}{N \times pos_{max}} \right\rangle_{left\ bank} \times \left\langle \frac{\sum_{n=1}^{N}(pos_{max} - (pos_n - pos_{n-1}))}{N \times pos_{max}} \right\rangle_{right\ bank} \tag{5}$$

where

$$pos_{max} = \langle max(pos_{N \in n}) - min(pos_{N \in n}) \rangle_{leaf\ bank} \tag{6}$$

The next characteristic is the aperture area variability (AAV), this represents the variation of the area relative to the maximum open area of the beam. The maximum open area of a beam is formed by taking the maximum positions (furthest away from the center) that leaves occupy during the irradiation of a full beam and combining all the maximum coordinates for each bank. To calculate the AAV of a segment, the distance between the left and right leaf is calculated for every leaf pair and summed up. This gets divided by the same calculation but then with the maximum leaf coordinates of the left and right leaf in a leaf pair throughout the beam.[35]

$$\mathrm{AAV} = \frac{\sum_{a=1}^{A} \langle pos_a \rangle_{left\ bank} - \langle pos_a \rangle_{right\ bank}}{\sum_{a=1}^{A} \langle \max(pos_a) \rangle_{left\ bank \in beam} - \langle \max(pos_a) \rangle_{right\ bank \in beam}} \tag{7}$$

The score starts at 1.0 for a normal rectangular open field and reduces the more complex a beam gets. The MCS can also be calculated for a full irradiation plan, for this the MCS of each beam is multiplied

with the weight of the monitor units of the beam relative to the total amount of monitor units in the plan, the MCS of the plan is then the summation of the results of this multiplication for every beam. [35]

2.4.2 Modulation index

The modulation index (MI) was first proposed by Webb in 2003. The MI uses intensity fluence maps to calculate transmission on a beam element (bixel) map. The modulation index calculation is done by comparing the deviation of contiguous bixels in a spectrum of the intensity fluence map to the standard deviation of the entire intensity fluence map. [33], [36], [37]

$$\mathrm{z(f)} = \left(\frac{1}{(\mathrm{n}-1)}\right) \mathrm{N(f;}\Delta_p > \mathrm{f}\sigma_I) \tag{8}$$

In this function, n is the number of sequential elements of bixel intensity I_p with p=1,2, 3, …, n. From these intensities, the average intensity and the standard deviation (σ_I) can be calculated. N is the amount of changes for which the Intensity difference between the bixel and an adjacent bixel Δ_p exceed a fraction f (f=0.01,0. 02, …,2) of the standard deviation σ_I. The z(f) function is a spectrum that decreases monotonically as the value of f increases. The value of the modulation MI is the area under the curve formed by the z(f) function and can therefore be written as an integral: $MI = \int_0^{0.5\sigma} z(f')df'$. The result of this integral will go up with increasing modulation of the beam. This is a widely used metric in the research for new complexity metrics, some modified versions have been made, but they all come down to this original version. [36]

2.4.3 Fluence map complexity

Another metric that uses fluence maps to calculate the complexity of an irradiation plan is the fluence map complexity (FMC), this compares the fluence in the bixels on a more local scale only taking into account the two lateral neighbours of a bixel. The fluence map complexity is a measure of smoothness of the fluence profile of the beam. When the fluence of a bixel is the same as the two lateral neighbours, the outcome will be zero the maximum value of the FMC metric is 1, this would represent a chequerboard pattern. A high fluence difference will have a lot of weight in the calculation of this metric because it gets squared before it is summed with the values of the other bixels. [33], [38]

$$\mathrm{FMC} = \frac{1}{\sum_j aj}\sqrt{\sum_j \left(a_j - \lambda_k \sum_{k\in N_j} a_k\right)^2} \tag{9}$$

The term a_j represents the fluence bixels that are being analysed, a_k stands for the neighbouring bixels of the a_j and λ_k is a weight assigned to each neighbouring bixel, in the case of two lateral neighbours, this is 0.5. [33], [38]

2.4.4 Small aperture score

The small aperture score (SAS) takes the distance between the leaf on the left and right side of a leaf pair as the defining parameter for its complexity score. A distance between the two leaves must be chosen as a criterion, if the distance between the leaves is shorter than the chosen criterion, the aperture between the leaves is considered as a "small aperture". This score is calculated for separate beams and multiplies the ratio of open leaf pairs below the chosen criterion to all the open leaf pairs not behind the jaws with the ratio of monitor units of the segment to the total amount of monitor units in the beam.[33]

$$\mathrm{SAS(x)}_{beam} = \sum_{i=1}^{I} \frac{N(x>a>0)_i}{N(a>0)_i} \times \frac{MU_i}{MU_{beam}} \tag{10}$$

In this formula I is the number of segments in the beam, x is the chosen criterion for the distance between a leaf pair. N is the number of MLC leaves that are not positioned under the jaws and a is the distance between opposing leaves of the aperture. [33]

2.4.5 Converted aperture metric

Like the small aperture score, the converted aperture metric (CAM) is based on the apertures in a beam. It measures the distance between leaves, not only parallel to the direction of the MLC leaves, but also perpendicular. The distance is measured every couple of millimetres, this distance is chosen to try and have a measurement for every MLC leaf pair in the parallel direction of the MLC leaves. To penalize short distances compared to large distances, a conversion function is used where $f(x)=1-e^{-x}$ this results in all the measured distances (d_i) having a value between 0 and 1, where 1 is noncomplex and given to all distances larger than 4cm. This is based on the assumption that problems with small fields are limited to fields smaller than 4cm x 4cm. Other than the measured distances, the equivalent square field size (a_{Eq}) of the MLC opening is used, this is the square root of the total area of the MLC opening. The conversion function used for the measured distances was also used for the equivalent square field size to penalize small areas [39], [40]

$$\text{Converted aperture} = 1 - \overline{f(d_i)} \times f(a_{Eq}) \tag{11}$$

Formula 11 calculates the converted aperture metric for one MLC opening of a beam segment, the CAM of the complete segment is the average of the CAM's of the control point. By subtracting the multiplication of the mean converted distance with the converted equivalent square field size from one, the converted aperture metric ranges between zero and one, where one is a complex MLC opening and zero a non-complex opening. [39]

2.4.6 Edge area metric

The edge area metric assumes that most dose differences between the calculated dose and measured dose come from the penumbra. To calculate this metric, two areas are defined. First, there is the edge region (R_1) that encompasses the penumbra at the edge of the open area. This area is defined as 5mm inside the field and 5mm behind the MLC leaves for the MLC edges that define the open area. The second area (R_2) is the open area of the beam not included in the edge area. The edge area metric is calculated by dividing the edge area by the sum of the edge area and the open area not enclosed in the edge area.[39]

$$\text{Edge area metric} = \frac{R_1}{R_1+R_2} \tag{12}$$

A simpler version of this metric exists as well, where the circumference of the aperture is divided by the area within the circumference[39]

2.4.7 Use of complexity metrics

There is still a lot of research being done to find new and better metrics. The metrics are mostly based on a few concepts of what leads to more complex plans: the MLC movement, small fields, the fluence maps and the administered monitor units. MU/Gy on its own is also a metric used to quantify the complexity, this is based on the observation that to achieve the same absorbed dose, more complex plans need to deliver a higher amount of monitor units.[33], [39]–[42]

The mentioned metrics have shown correlation with gamma passing rates but show mixed results for the purpose of reducing the workload of the physics department of hospitals. This can be achieved by setting threshold criteria for the complexity. If a plan meets the criteria (low enough complexity) the pre-treatment quality assurance test is not necessary anymore. Or a threshold for high complexity scores indicating which plans are best to recalculate with a lower complexity because the chance that it would fail the pre-treatment QA is significant if the complexity is above that threshold. To do this with a high enough certainty, thresholds are chosen at complexity values that give low false positive rates and high true positive rates. [33], [39]–[42]

3 Method and materials

3.1 Complexity metric

In the literature study a couple of complexity metrics were discussed. In this research, a different metric will be used: the edge metric developed by Younge et al from the University of Michigan. The choice for this metric was because it had shown promising results in the research that was performed by Younge et al to predict the outcome. And because in their research a script was used to do the calculations which was the plan to do in this research as well. This gave a good foundation to start the research in this project to try and implement it in the LOC. The metric used in the research by Younge et al is an aperture and edge-based complexity metric. It was derived from an earlier research project where they developed it as an edge penalty to implement in the in-house treatment planning system to reduce abundant complexity of VMAT plans calculated with inverse optimization. Without the developed penalty the errors between calculated and measured dose concentrated around the edges of the apertures of the calculated plans. This can be explained by the difficulty that the calculation algorithms have with calculating the penumbra of a beam correctly. The penumbra is a region with a high dose gradient, small displacements, or miscalculations in the width of it can result in big dose differences. If the aperture area is small, a deviation in the penumbra will have a bigger effect on the dose delivered by the aperture compared to when it is a large area. [1], [2]

$$P = C \sum_{i=1}^{N} W_i \times \frac{C_1 x_i + C_2 y_i}{A_i} \tag{13}$$

The original metric used for the penalisation of the amount of edge in an aperture is given above. It is the sum for N apertures with i as the index for the control point aperture that is calculated. W_i is the weight of the aperture in the beam, this is the contribution to the total dose (monitor units) of the beam, C_1 and C_2 are scaling factors to give more importance to either the edges of the leaf ends (x_i) or the leaf sides (y_i), A_i is the total open area of a control point that is not covered by the jaws of the linac. [1], [2]

C is a global scaling factor used for the penalization that defines the importance for the inverse optimization relative to the importance of the dose-related cost function. This determines which cost function the calculation algorithm should consider more. This is not used when calculating the edge metric for the comparison with gamma passing rates to single out plans with a higher risk of failing the quality assurance tests. [1], [2]

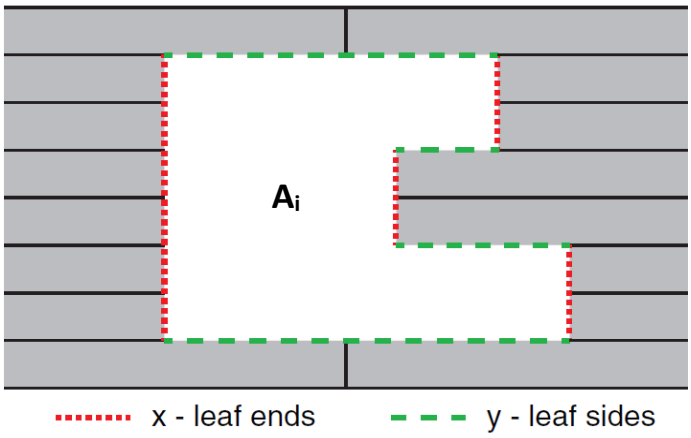

Figure 3 Example of an aperture of a control point with the x and y edge parameters indicated.[33]

$$\mathrm{EM} = \frac{1}{MU}\sum_{i=1}^{N} MU_i \times \frac{y_i}{A_i} \quad (14)$$

Formula 14 is the actual complexity metric used to determine the complexity for the purpose of this research. It is also the formula used by Younge et al when they researched the possibility to use the penalization metric for the purpose of predicting the deliverability of VMAT irradiation plans. The first difference with formula 13 is the global scaling factor that is gone and the weighting factor that is split up in two parts, the division by the total number of monitor units of a beam (MU) is done after the summation. The amount of monitor units given to a patient with one segment is inside the summation sign where it is multiplied with the edge metric of a single control point. This is to give higher importance to the control points that contribute the most to the delivery of the monitor units to the patient. This is important since more monitor units itself is also an indicator of more complex plans or segments. [1], [2]

Another difference is the fact that the scaling factors C_1 and C_2 are not in the formula and so is the x_i parameter that indicated the sum of the length of the leaf ends surrounding the open area. This is because C_1 is chosen as 0 and C_2 as 1, resulting in the omission of the x_i leaf end parameter. The scaling factors are chosen this way because the leaves are either open, or closed, there is no in between so the only way to change the x_i parameter was by opening or closing a leaf pair, but open leaf pairs are no indicator of plan complexity or irregularities, so it would not make the edge metric perform better. The edge metric starts at 0 which is not complex the higher the number for complexity, the higher the complexity is, the unit of the metric is mm^{-1}.[1], [2]

The previous formula (14) calculates the edge metric for one beam or arc, to calculate the metric for a full plan, y_i/A_i is replaced by the complexity of one beam or arc in the treatment plan. This gets multiplied by the amount of monitor units of that beam or arc. This calculation is performed for all the beams or arcs in the plan. Those values are added up and then divided by the total number of MU of the plan. [1], [2]

$$\mathrm{EM}_{plan} = \frac{1}{MU_{plan}}\sum_{i=1}^{N} MU_{arc\ or\ beam_i} \times EM_{arc\ or\ beam_i} \quad (15)$$

This metric shows similarities with the small area score (2.4.4) and the edge area metric (2.4.6) since they are based on the same principles to determine the complexity of a plan: the penumbra on the edge of the MLC leaves that can induce relatively big calculation errors. But there are some differences: the small area score identifies leaf pairs with small distances between the left and right leaves, but it does not identify if the aperture between leaf pairs is connected with leaf pairs of the adjoining leaf pairs. The edge metric that is used in this research does differentiate between the small apertures between leaf pairs being adjoint or separately scattered around the control point. The edge metric shows the most similarity with the edge area metric and more specifically with the robust version: the circumference divided by the area. The division of y_i by A_i is only different from the circumference divided by the area by omitting the top leaf ends from the circumference. While in the edge area metric, it is not just the length of open sides, but an area around the open sides that is taken into account. [1], [2]

3.2 Software and appliances

3.2.1 Planning, delivery, and QA of treatment plans

All the plans that were analysed in this research were created with Eclipse™. This is the treatment planning system created by Varian (Varian Medical Systems, Palo Alto, CA). The VMAT plans were made with RapidArc™. Eclipse is part of ARIA, an oncology information system by Varian. ARIA also consists of quality assurance applications. The portal dosimetry application in ARIA is used to analyse the differences between the calculated plans and the pre-treatment QA verification plans delivered onto the PortalVision™ MV imaging system. This gives gamma passing rates in return (see 2.3.2 on how this is calculated). These passing rates can be calculated based on absolute or relative normalization. The absolute method compares calibrated units (CU), these calibrated units are based on the pixel values of the EPID and constructed by calibration measurements. At the Limburgs Oncologisch Centrum (LOC) in Hasselt, the criteria for an irradiation plan to pass the patient specific pre-treatment QA is a passing rate of 97% for 3% dose difference and 3mm distance to agreement criteria for the global gamma index.

All plans were irradiated with 2 machines, first of all there is the Varian Clinac™ with a Millennium 120 multileaf collimator (MLC 120). This MLC consists of 120 leaves, with each bank having 60. The outer 10cm at both sides of the MLC bank consist of MLC leaves with a width of 10mm while the inner 20cm consists of leaves that are 5mm wide. Both banks together form a field of 40cm x 40cm. The second type of linac used at this institution is the Varian TrueBeam™ linear accelerator equipped with a HD120 multileaf collimator. This also has 120 leaves, but the outer 14cm of the bank are leaves with a leaf width of 5mm, the inner 8 cm of leaves are 2.5mm wide, forming a field of 40cm x 22cm. Both types of linacs are equipped with a PortalVision™ electronic portal imaging device.

3.2.2 Development and implementation of the complexity script

Doing all the calculations of the metric by hand would take hours to do for even one arc, application scripts are a useful tool to help with the calculation of complexity metrics. Varian has build a tool to work with scripts in the treatment planning system to aid users in this process. This is the Eclipse Scripting Application Programming Interface (ESAPI), the scripts for this tool have to be written in the Microsoft C#, this is a C++ based programming language. To write the script Microsoft Visual Studio 2017 was used.

To start this project, the first task was to learn how to script, how everything in C# worked to be able to develop a script that could carry out these calculations. Younge et al. shared the script that they used on Github so it could be used by others to try and implement it and do more research on it. Because it was made in an older version it was not possible to just implement it as they made it. The code for this project was therefore based on the code from their research but tailored to the needs of this research and adapted to work with the current software used at the LOC in Hasselt.

Two versions of the script were made, the backend of how it calculates the metric is the same, but one version can be used to calculate complexity metrics for whole lists of patient ID's to perform analysis of the complexity for multiple patients for research purposes like this Master's thesis. The other version is a plug-in that can be implemented in the daily working routine of the physics department of the LOC. It can be used to calculate the complexity during the planning stage of irradiation plans, this allows the complexity value to be compared with the average complexity for a specific type of plan and if possible with threshold values to determine if pre-treatment QA verification tests on the linac are necessary or that the complexity has to be lowered.

The script opens a patient or uses a patient that is open in the treatment planning system, reads out the coordinates of the MLC leaves in both banks and the position of the Jaws. It calculates the difference in coordinates of the top edge of a leaf compared to the top edge of the adjacent leaf of the bank. If the coordinates of the top edge of a leaf overlaps with the coordinates of where the Jaw is, indicating the MLC leaf is behind the jaw, it will return 0 as value. All these distances of both banks are summed up in parameter y_i of formula 14. By calculating the area between open leaf pairs and summing up these area's that are not behind the jaws, parameter A_i is computed. To correctly get all the coordinates and calculate the aperture areas between open leaf pairs, the width of the leaves in the MLC arrays of the Clinac™® and TrueBeam™ is hardcoded inside the code of the script, based on the linac an irradiation plan is made for, the script will choose the right composition of MLC leaves. It is important to note that the coordinates are attributed relative to the centre of the MLC-field, resulting in negative coordinates depending on the position of the MLC leaf in the field.

The script calculates the complexity for all plans and the associated arcs, beams or fields attached to the chosen treatment diagnosis code in the patient course. This way multiple plans can be made and they can be compared to each other to see which one has the lowest complexity. In figure 4 this would result in an extra part Plan: prostaat and then the results of the calculation. The standalone version (figure 5) will create .txt files with a list of data:

- Patient ID
- Plan ID
- Complexity metric of the plan
- Treatment unit ID
- Beam ID
- Control points of the beam
- Amount of MU's of the beam
- Complexity metric of the beam
- Gantry direction

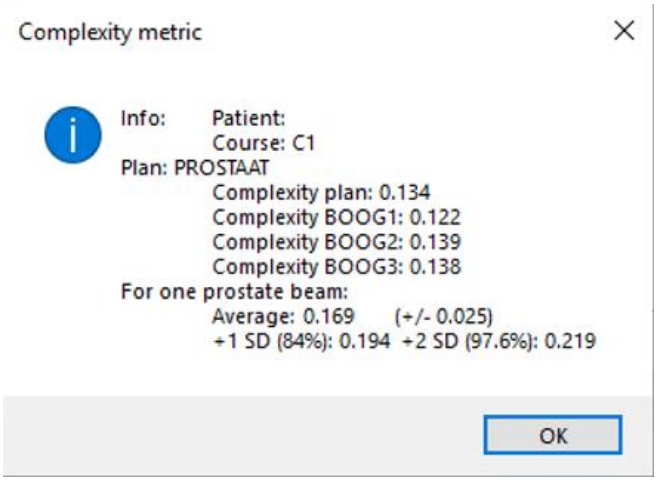

Figure 4 Example of the plug-in output window for a prostate plan with first the complexity of the plan, then the complexity of each arc and then the average complexity and the standard deviation for prostate arcs.

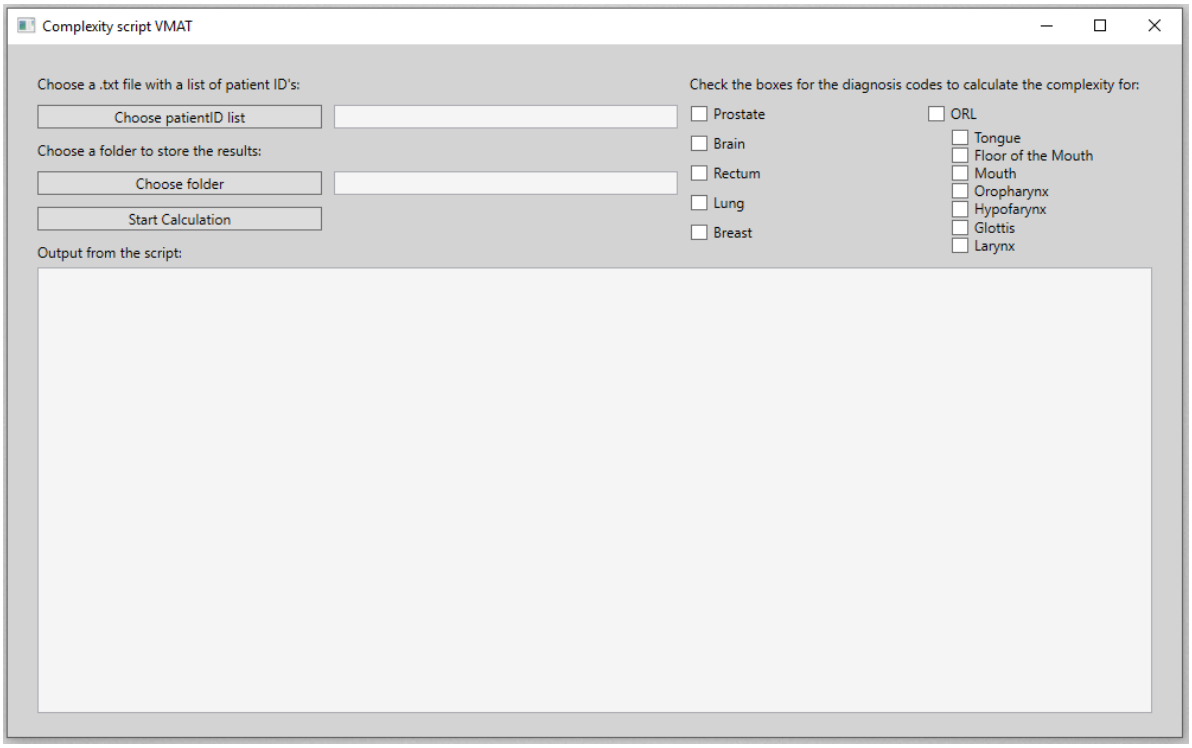

Figure 5 Screenshot of the standalone version of the script used to analyse lists of patient ID's.

3.3 Data analysis

3.3.1 Using of the script

As all the calculations are carried out by the script, it is important to make sure that the calculations are performed correctly. This can be done in Microsoft visual basic when running the code. By adding breakpoints and following the execution of the code it is possible to follow all the data through the script and verify if the correct data goes through every part of the calculation.

Other than that, it was also verified if the results are as expected. To do this irradiation plans with different degrees of modulation were made for a prostate VMAT plan. 4 different plans were made: one without modulation where the arc is just an open beam where the MLC leaves from the shape of the tumour volume and three other plans with modulation settings ranging from light to high. By allowing the treatment planning system to modulate a plan more, the MLC leaves will move around more resulting in less uniform fields. This should result in a higher number for the complexity metric if the calculation is performed correctly. These plans were irradiated onto a 2D array of ionisation chambers, the MatriXX universal detector array was used for this verification. (Iba, Louvain-La-Neuve, BE).

3.3.2 Correlation

The Pearson correlation coefficient will be used to quantify the correlation between different parameters. This value ranges between -1 and 1, -1 stands for a perfect descending, negative linear correlation, meaning that they move in opposite directions of each other, if one parameter goes up, the other will go down and vice versa. 1 is a perfect ascending, positive correlation which means both parameters increase or decrease at the same rate. An r value of 0 means that there is no correlation between the parameters that are being researched. The calculation of these values is done with Microsoft® Excel. The correlations of the metric, MU and gamma passing rates will be examined.[43]

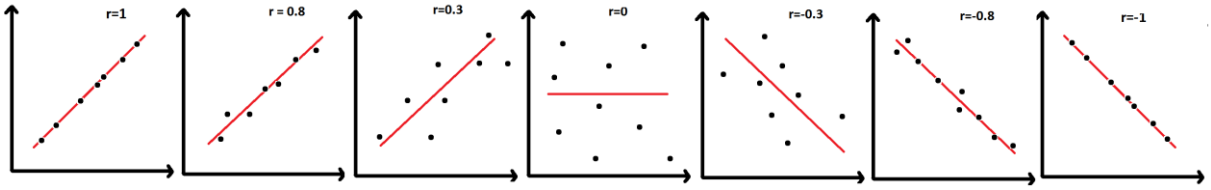

Figure 6 Example of correlation graphics and R-values, starting at a perfect positive correlation on the right (r=1), no correlation in the middle graph (r=0) and perfect negative correlation on the left (r=-1) [35]

3.3.3 Receiver operating characteristic

With the determination of thresholds to predict the outcome of the pre-treatment QA verification being the original intent of this research, a method is needed to analyse the diagnostic ability of the complexity metric to separate irradiation plans that will pass the pre-treatment QA verification and which will fail the verification. The hypothesis that will be investigated is that a high value of the complexity metric will result in low gamma passing rates leading to a failed pre-treatment QA verification and plans or beams with a low complexity metric value will accord to plans with high gamma passing rates and thus passing the verification. [2], [44]

For this purpose, the receiver operating characteristic (ROC) and the corresponding area under curve (AUC) are widely used techniques. In this method the true positive rate (TPR) and false positive rate (FPR) for different threshold values are plotted in a graph, this is called the ROC-curve, the area under that ROC-curve is as the name says the AUC. [2], [44]

Table 1 The conditions for plans to be categorized as True Positives (TP), False Positives (FP, False Negatives (FN) and True Negatives (TN)

		Actual value	
		γ < Passing Rate	γ > Passing Rate
Predicted condition	EM > Threshold	True Positive (TP)	False Positive (FP)
	EM < Threshold	False Negative (FN)	True Negative (TN)

$$\text{TPR} = \frac{True\ Positives\ (TP)}{True\ Positieves\ (TP) + False\ Negatives\ (FN)} = \frac{True\ Positives\ (TP)}{All\ positives\ for\ the\ actual\ value} \tag{16}$$

$$\text{FPR} = \frac{False\ Positives\ (FP)}{False\ Positieves\ (FP) + False\ Positives\ (FP)} = \frac{True\ Positives\ (TP)}{All\ negatives\ for\ the\ actual\ value} \tag{17}$$

Table 1 shows how the plans, arcs and beams will be distributed in four groups depending on the gamma passing rate and the edge metric. To get a true positive value in this research, the gamma index (γ) has to be below the set gamma passing rate threshold for the chosen constraints (e.g. 97% for 3%|3mm) while the complexity metric is above the threshold for which the TPR and FPR are being calculated with formulas 16 and 17. [2], [44]

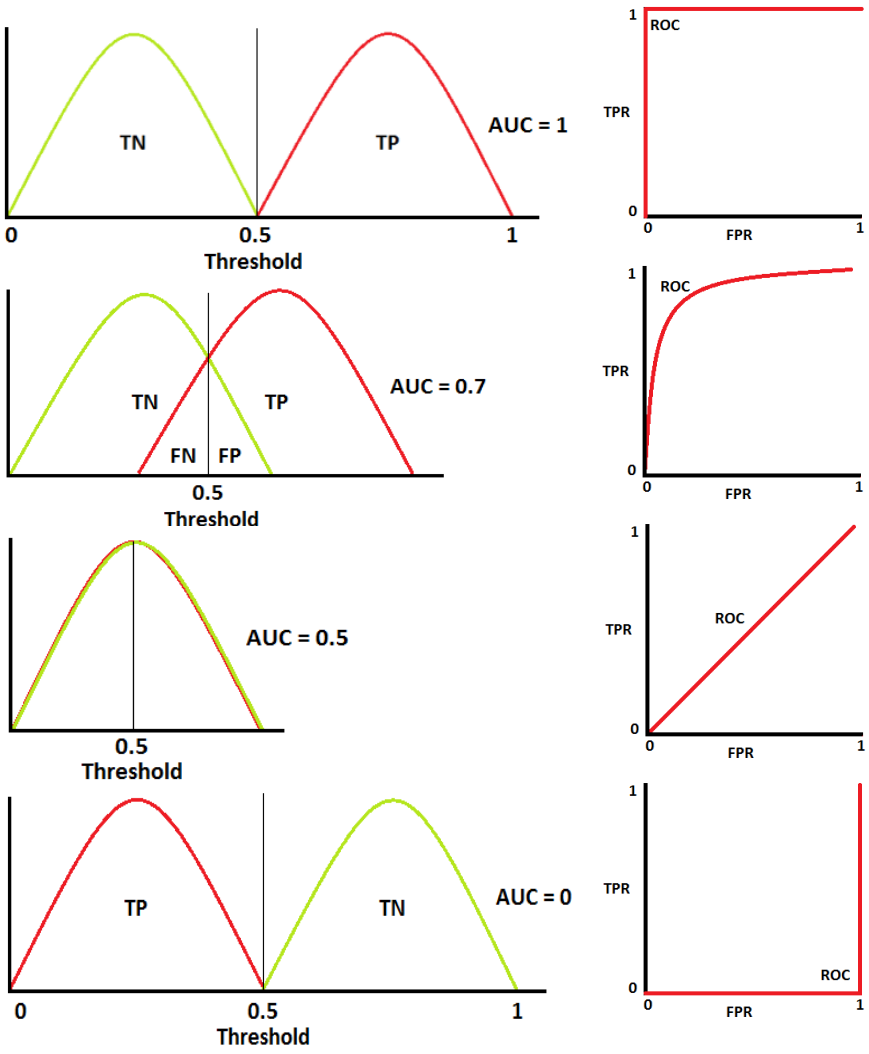

Figure 7 Graphic example of how ROC and AUC work [35]

A graphical example of how the ROC and AUC work is depicted in figure 7. Each data point that forms the red ROC-curve represents a threshold and every threshold comes with a true positive and false positive rate. It visualizes how correctly a certain parameter, in this research the edge metric, can differentiate the 2 groups of values: the plans that fail the pre-treatment QA (TP) and the plans that succeed (TN). If a parameter can separate them perfectly, the first graphs is produced. The area under the curve is one, this is the probability that the chosen parameter can correctly differentiate random values in the correct group, which is 100% when this is done perfectly. If there is an overlap like the second set of graphs show, this percentage will be lower, and the curve will not be a perfect right angle. This means that all the values in that overlap will not be guaranteed to be put in the correct group lowering the accuracy of the chosen parameter.

If both sets of values overlap completely, the ROC will form a straight line and the AUC will be 0.5, this means that there is a fifty-fifty chance it will correctly predict the outcome. This is a situation that results in no predictability of the chosen parameter to differentiate a group of values. The last set of graphs shows what happens if it predicts all values wrong, resulting in the mirror image of the first example. If this happens, the opposite of the hypothesis would be true.

3.3.4 Which data to analyse

The plans that will be analysed during this research were all delivered during pre-treatment QA on either a TrueBeam™ or Clinac™ at the LOC. With another script, a list of the last ±300 patients was generated for the six most prevalent treatment sites that are being treated with intensity modulated plans at the LOC:

- prostate,
- rectum,
- brain,
- lung,
- breast and glans,
- otorhinolaryngology (ORL).

All the generated lists of patient numbers are put into the script that calculates the edge metric returning a file with all the parameters listed in part 3.2.2. All these lists are also run through a second script that returns the gamma passing rates that were used to validate the irradiation plan, the separate arcs and the beams. This means that this contains both values of the absolute and relative method to calculate the gamma passing rate. To be able to correctly compare the edge metric values to the gamma passing rates, the passing rates have to be generated for the same conditions. Getting these passing rates, requires every patient file to be opened manually, the parameters changed to the preferred values for the evaluation and writing them down in a spreadsheet..

4 Results

4.1 Control of the script

After making sure the script worked correctly by looking at the dataflow in the execution, the following plans (table 2) were generated in the Eclipse treatment planning software. All the plans were made for the same patient with modulation set at no modulation, light-, medium- and high modulation in the RapidArc™ module The measurements of the gamma passing rate were done with the Matrixx 2D Array. The results are visible in table 2.

Table 2 Complexity and gamma passing rates for plans with four different amounts of complexity.

Criteria	Complexity (mm^{-1})	γ (3%|3mm) (%)	γ (3%|1mm) (%)
Open beam	0.023	100	100
Low modulation	0.065	99.8	99.1
Medium modulation	0.097	99.3	98.1
High modulation	0.253	97.2	91.9

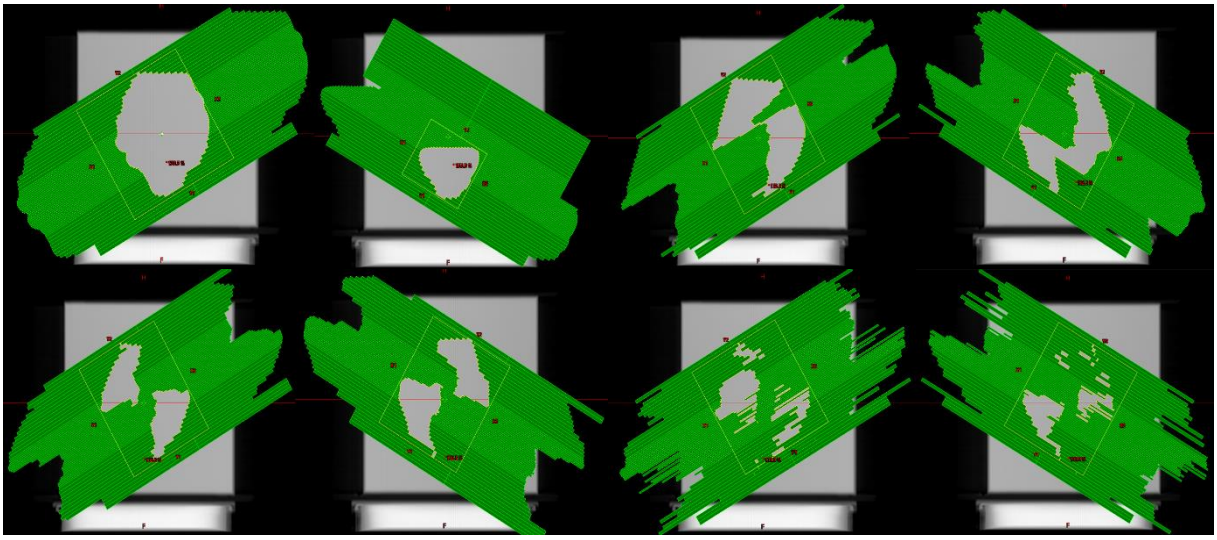

Figure 8 MLC positions at the same angle in the arc for four plans with different amounts of complexity of the same patient. Top left: no modulation, top right: low modulation, bottom left: medium modulation, bottom right: high modulation.

4.2 Correlation between the measured variables

4.2.1 Correlations for treatment plans for Clinac™ accelerators

The data that follow from here on are most complete for the VMAT prostate plans delivered with a Clinac™. For other treatment sites and linacs it was planned to do the same analysis as will be shown for the VMAT prostate plans on the Clinac™, but due to the COVID-19 outbreak and the lockdown that followed the collection of gamma passing rates was not possible anymore. For the six most prevalent treatment sites complexity values were calculated, these will be investigated by looking at their averages. The values in table 3 and figures 9 to 13 are from the VMAT treatment plans for the Clinac™, 118 patient irradiation plans were investigated containing 236 VMAT arcs.

Table 3 R-values to determine correlation between MU, EM and gamma passing rates.

Researched correlation	R-value
MU vs EM (plans)	0.559
MU vs EM (arcs)	0.502
Absolute γ (3%|3mm) vs EM	0.163
Relative γ (3%|1mm) vs EM	-0.252
Relative γ (2%|1mm) vs EM	-0.289

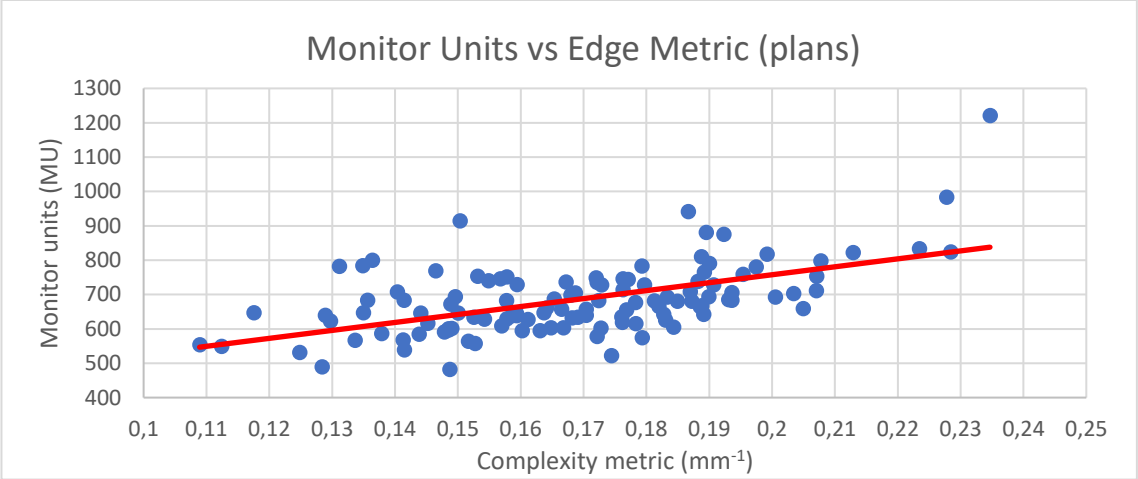

Figure 10 Correlation between monitor units and the complexity metric of full Clinac plans.

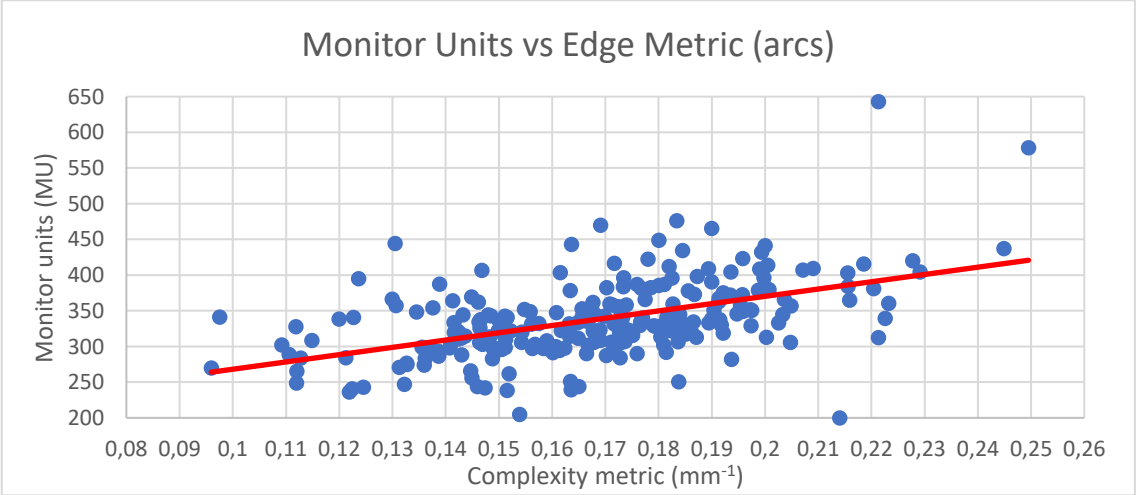

Figure 9 Correlation between monitor units and the complexity metric of single Clinac arcs.

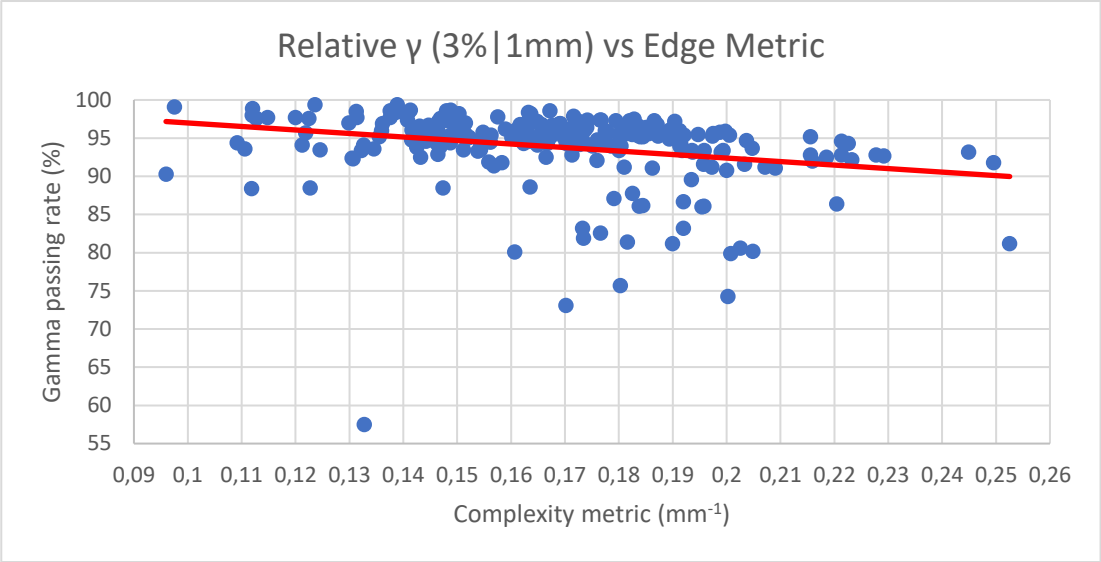

Figure 11 Correlation between relative gamma passing rates (3%|1mm) and the complexity metric (Clinac).

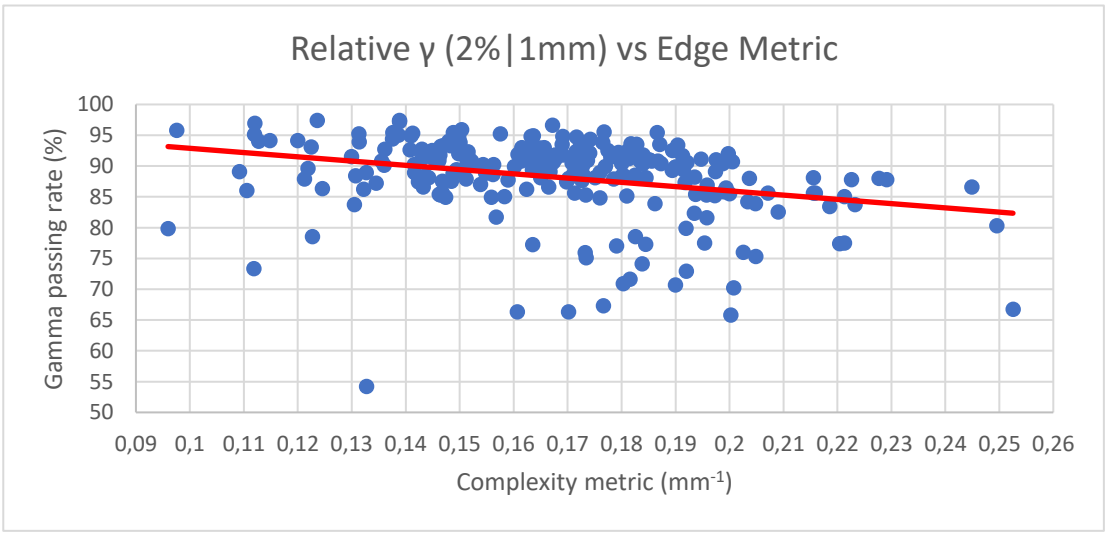

Figure 12 Correlation between relative gamma passing rates (2%|1mm) and the complexity metric. (Clinac)

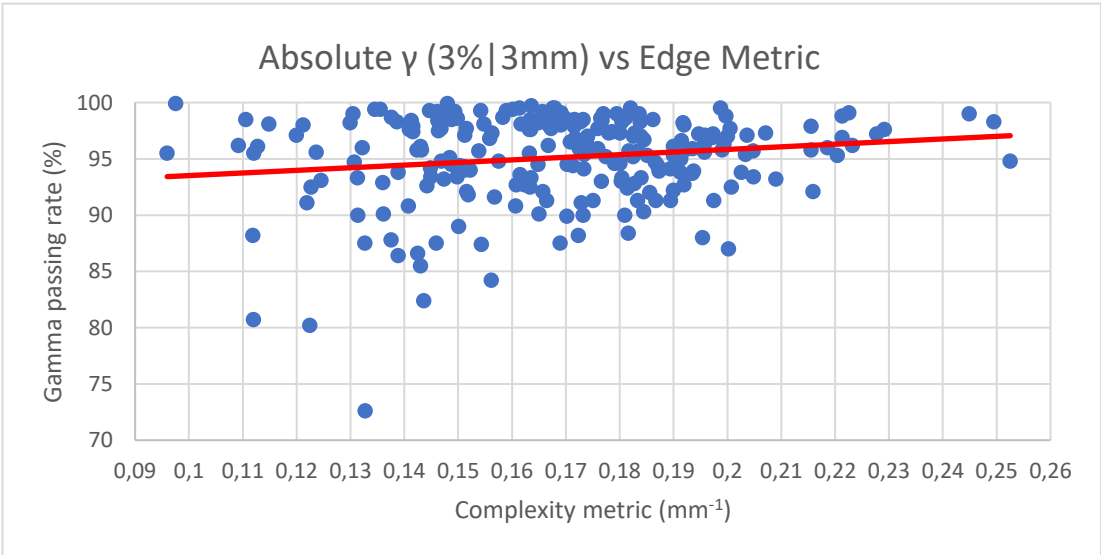

Figure 13 Correlation between absolute gamma passing rates (3%|3mm) and the complexity metric. (Clinac)

All the gamma analyses that were investigated had a global passing rate of 10%. For the absolute 3% dose difference 3mm distance to agreement gamma test, 59.7% of the arcs (141 out of 236) scored below the 97% threshold. (In paragraphs 4.3.1 and 5.2 is explained what causes this and how this is handled.) In figure 14 the result of a pre-treatment QA test is visualised with beam profiles generated by the portal dosimetry software in Eclipse. Both were taken in the same place in the beam, the generated dose map is one map for a full beam, it is not separated in separate control points. The blue and green curves represent the measured plan by the EPID (BOOG1-4_1_3). The red and yellow curves represent the calculated plan, this is how it predicts it will be delivered.

The top graph in figure 14 represents the absolute normalization, with the calibrated units on the y-axis and the distance from the isocenter of the beam on the x-axis. The bottom graph gives the curves for relative normalization with the minimize difference technique that is used for all relative normalizations in this paper. The portal dosimetry software tries to match the dose maps of the calculated and measured as best as possible on top of each other and gives a passing rate based on this normalization. This takes away the factor of the calibration of the EPID generally resulting in higher passing rates.

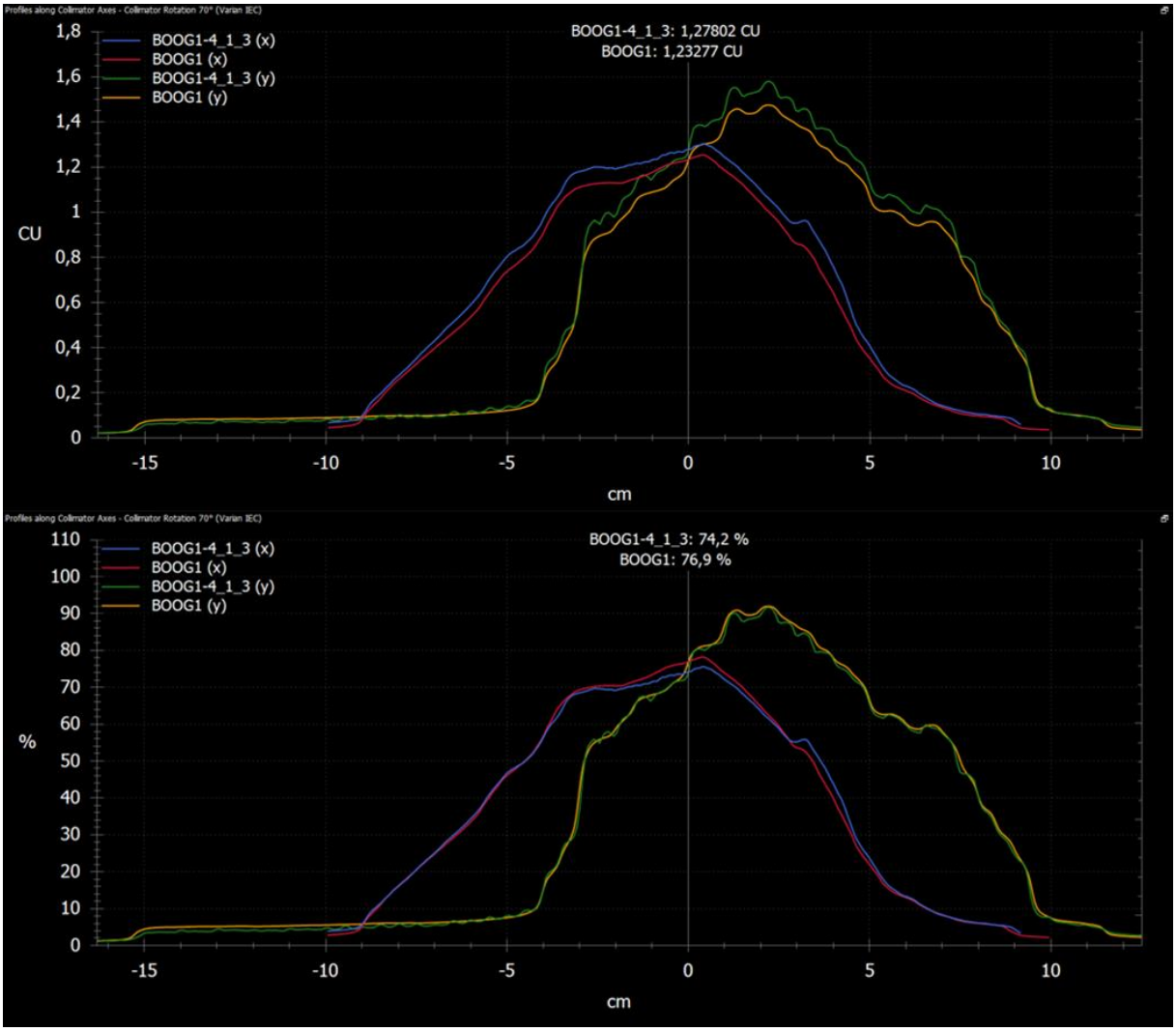

Figure 14 Example of gamma passing rate with absolute normalization (top image) and relative normalization with minimize difference (bottom image)

4.2.2 Correlations for treatment plans for TrueBeam™ accelerators

For the VMAT arcs treated with the TrueBeam™, the first set of gamma passing rates for the absolute test with 3% dose difference and 3mm DTA was completed. For this data, the same plots were made as in 4.2.1 for the Clinac™-plans. It can be seen that most data points are close to 100%, 88.4% (275 of 311) of the arcs have a gamma passing rate of 97% or higher, 85.9% of the passing rates are above 98% and when the passing rate threshold is set at 99%, 76,2% still pass the QA. When looking at the total amount of arcs that failed the pre-treatment QA (also including relative test 3% 3mm test when the absolute test failed), only 4.8% (15 of 311 arcs) failed the QA.

Table 3 R-Values TrueBeam™ plans

Researched correlation	R-value
MU vs EM (plans)	0.531
MU vs EM (arcs)	0.530
Absolute γ (3%|3mm) vs EM	0.384

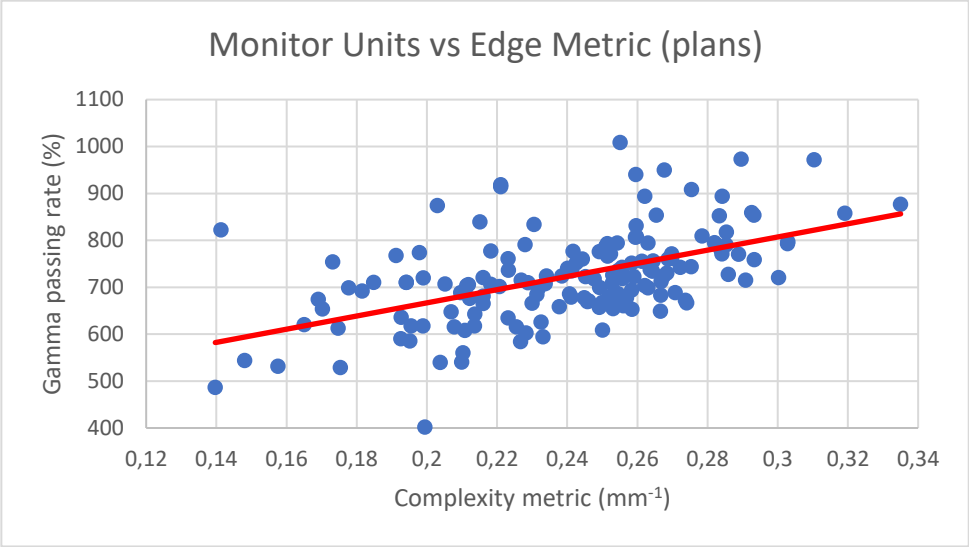

Figure 15 Correlation between monitor units and the complexity metric of full TrueBeam™ plans.

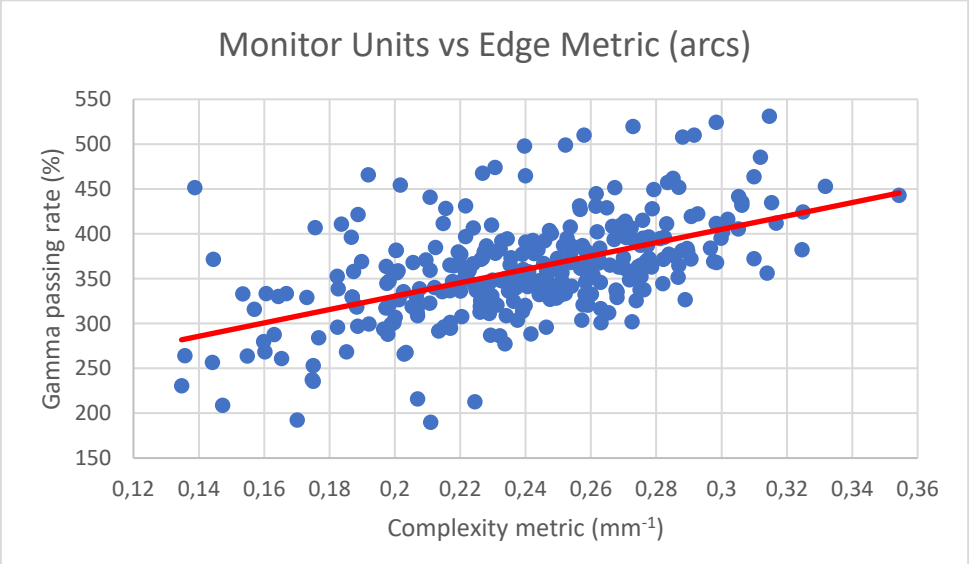

Figure 16 Correlation between monitor units and the complexity metric of single TrueBeam™ arcs.

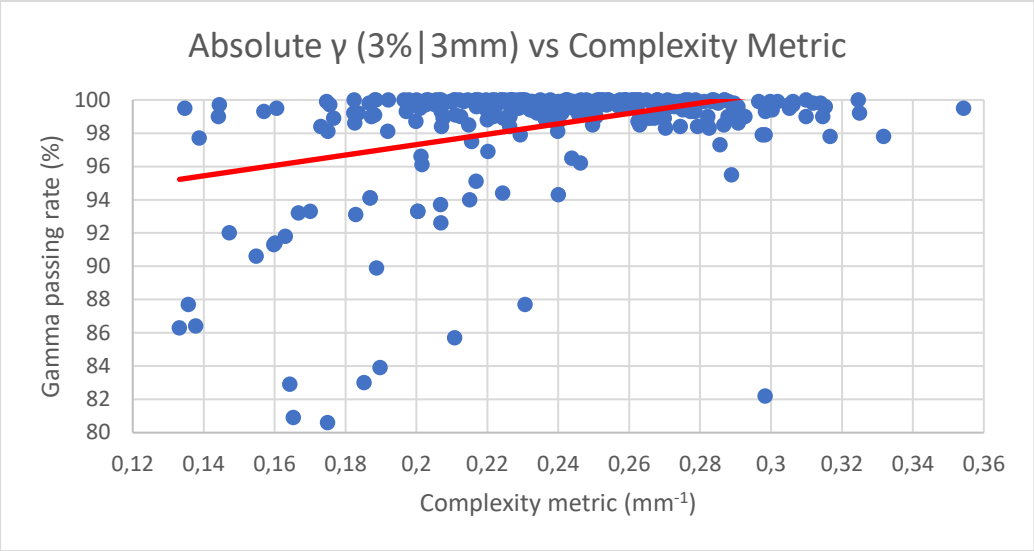

Figure 17 Correlation between absolute gamma passing rates (3%|3mm) and the complexity metric. (Clinac)

4.3 Results of the ROC-curves

4.3.1 ROC-curves for treatment plans for Clinac™ accelerators

The following graphs show the receiver operating curves (ROC-curves) for the studied gamma passing rates. The passing rates are the same as the ones used in 4.2 for the correlation between the passing rates and the complexity. Table 4 shows the values for the area under the curve for the three constructed ROC-curves. The passing rate of the absolute gamma test had to be above 97% as it is used in daily practice at the LOC. For the relative 3% dose difference and 1mm distance to agreement criteria, 95% of the control points had to comply to those criteria, for the 2% and 1mm relative test, this was dropped to 90%, the lowest suggested value by the NCS [15]

In practice at the LOC, the absolute gamma passing rate with 3% DD and 3mm DTA was used to do the first control of the pre-treatment verification. If the gamma passing rate is below the wanted 97%, arcs are often verified with a relative method (not only the minimize difference method) to see if it will reach the 97% passing rate or not. Arcs that fail QA on the absolute dose verification are sometimes still passed based on a relative dose verification. This is what is displayed on the fourth ROC-curve where the passing criterion is not a passing rate but the fact if a plan actually passed on either the absolute or a relative test, this is not always the minimize difference normalization that was used in this thesis to calculate the relative gamma scores. For some plans the criteria were also upped from 3% DD and 3mm DTA to 3.5% DD and 3.5mm DTA. This resulted in a reduction of 141 failed arcs with the absolute gamma passing rate test to 37 failed plans with the combined results.

Table 4 Area under curve for the different ROC-curves

Used Gamma test	Area under curve (AUC)
Absolute (3%|3mm)	0.565
Relative (3%|1mm)	0.685
Relative (2%|1mm)	0.682
Passed/Failed	0.575

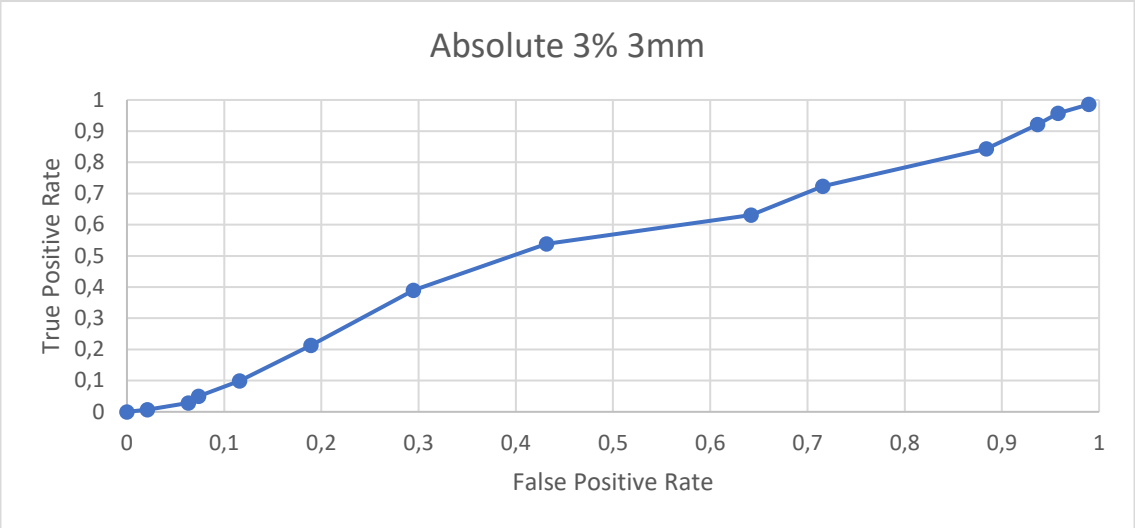

Figure 18 ROC-curve for the absolute gamma test with 3% DD and 3mm DTA (Clinac)

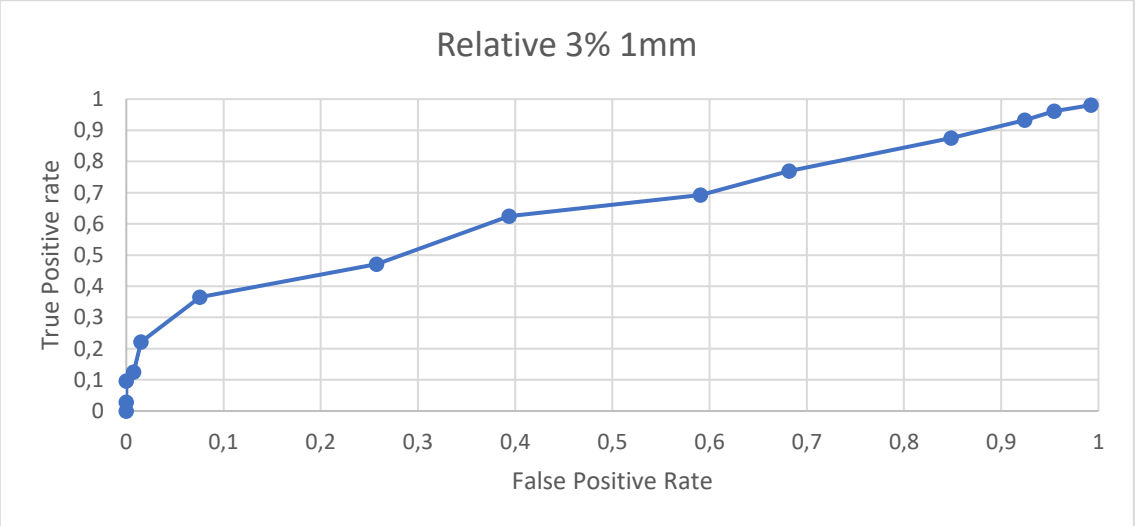

Figure 19 ROC curve for the relative gamma test with 3% DD and 1mm DTA (Clinac)

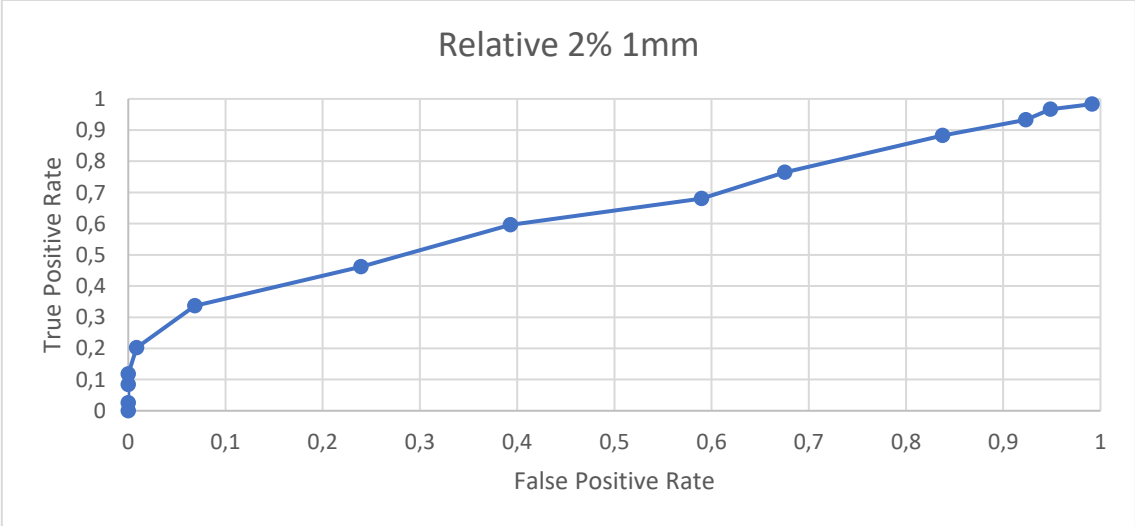

Figure 20 ROC-curve for the relative gamma test with 2% DD and 1mm DTA (CLinac)

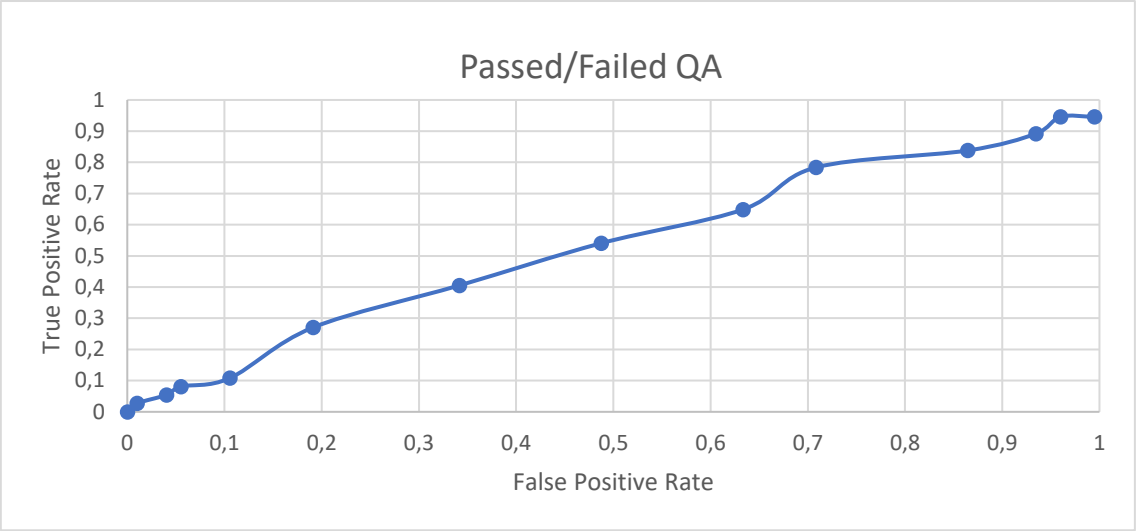

Figure 21 Passed/Failed QA absolute and relative combined (Clinac)

The last graph shows the number of plans that have a gamma passing rate higher than 95% for the 3% DD and 1% DTA criteria. The arcs are separated into groups with a difference of 0.01mm^{-1} between each group. This graph is a recreation of the example in 3.3.3. Once the complexity is above 0.190mm^{-1} more arcs fail the gamma test than succeed and a distinction between both curves can be made. There is a big overlap of the two curves, not a perfect separation as the first example at 3.3.3.

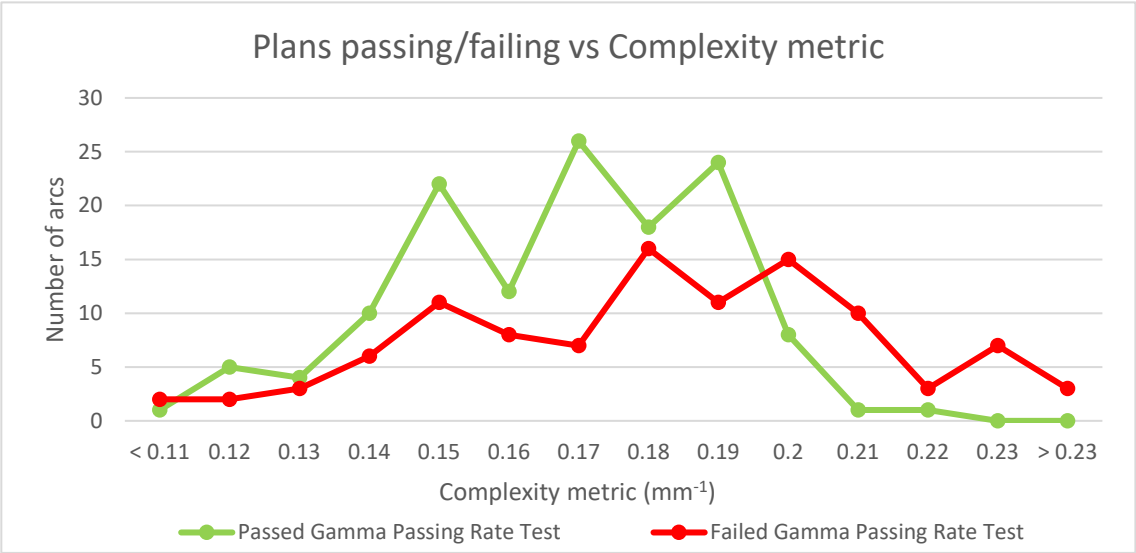

Figure 22 Number of arcs that pass (green) or fail (red) the relative 3% 1mm gamma test (Clinac)

4.3.2 ROC-curves for treatment plans for TrueBeam™ accelerators

For the TrueBeam™ plans it was only possible to make a ROC-curve for the absolute gamma test results with 3% dose difference and 3mm distance to agreement. Due to the Covid-19 crisis further gathering of gamma passing rates for different relative constraints wasn't possible. The area under curve for this ROC analysis is 0.193, this is because as seen in the scatter chart that put the complexity metric against the gamma passing rates in chapter 4.2.2, only few plans failed the gamma test with absolute normalization and most of the failed plans occurred in arcs with a low complexity score. To draw conclusions from this graph, the results from the relative gamma index tests are needed to compare them with the results from the Clinac™ accelerator treatment plans.

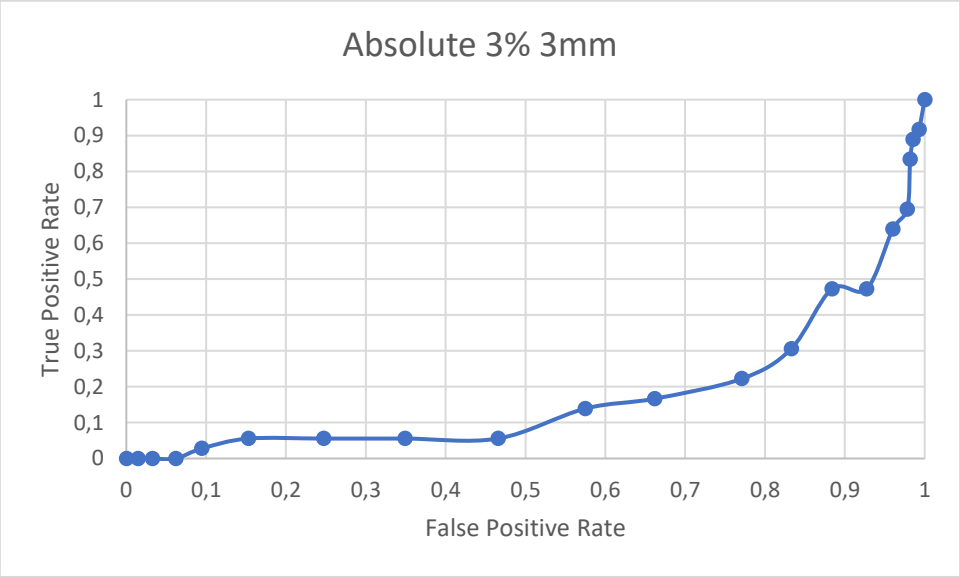

Figure 23 ROC-curve for the absolute gamma test with 3% DD and 3mm DTA (TrueBeam™)

4.4 Average complexity for different treatment sites and methods

Table 5 below this paragraph contains the average complexity of arcs or beams for the six most prevalent treatment sites that are treated with intensity modulated irradiation plans. The averages were not only separated for the different treatment sites, but also the treatment unit the irradiation plans were made for and different treatment techniques. For breast irradiations, the arcs were separated between those used for the first part of the treatment and the ones used for boost irradiations. The arcs for the first part were mostly 66 control point arcs, those for the boost irradiations contained 114 control points. The average amount of monitor units for the boost arcs was 499.77 MU (±104.27), while it was 98.71 MU (±46.88) for the normal VMAT arcs. That is because a part of the dose in the first part of treatment is given with static beams, while in boost treatments the full dose is given with an arc. The analysis for ORL irradiations on the Clinac™ was also subdivided into the number of control points used in an arc to investigate whether this affects the complexity metric.

Table 5 Averages for six most prevalent treated treatment sites with intensity modulated plans for different treatment units and treatment techniques

Treatment site	Treatment Unit	Complexity Metric (mm^{-1})	Standard deviation (mm^{-1})	Treatment method
Prostate	Clinac™	0.168	0.028	VMAT
	TrueBeam™	0.239	0.040	VMAT
Rectum	Clinac™	0.119	0.021	VMAT
	TrueBeam™	0.155	0.027	VMAT
Brain	Clinac™	0.091	0.038	IMRT
		0.108	0.041	VMAT
	TrueBeam™	0.084	0.023	IMRT
		0.150	0.060	VMAT
		0.341	0.077	SRS
Lung	Clinac™	0.134	0.033	VMAT
		0.170	0.038	SBRT
	TrueBeam™	0.204	0.060	VMAT
		0.226	0.055	SBRT
Breast	Clinac™	0.098	0.040	IMRT
		0.114	0.029	VMAT
		0.141	0.021	VMAT BST
ORL	Clinac™	0.127	0.025	VMAT
		0.128	0.022	VMAT 98 CP
		0.128	0.22	VMAT 114 CP
		0.126	0.027	VMAT 178 CP
	TrueBeam™	0.164	0.033	VMAT

5 Discussion

5.1 The script

Both scripts are working as expected and they can be used at the LOC for further research. Currently the plug-in will only be used to display the averages and standard deviations for the different VMAT RapidArc™ plans. Based on those averages and standard deviations, the first (84% of the plans have a lower complexity) and second standard deviation (97.6% of the plans have a lower complexity) from the average are calculated. These three values can be used to try and reduce the complexity of plans when planning RapidArc™ irradiation plans in the future. This is possible because the metric gives the physicists a quantitative number to define the complexity to refer to when planning.

The results of the plans that were specifically made to mimic no-, light-, medium- and high modulation gave the results that were expected based on the way the edge metric works. With only the high modulated plan resulting in a complexity above the average complexity for TrueBeam™ plans and a gamma passing rate close to the limit of 97% for 3%/3mm. The visual representation of MLC positions also confirms what kind of composition gives high complexities: the fields with a lot of small apertures and very irregular movements of MLC leaves. This also suggests that most TrueBeam™ plans for prostate at the LOC are planned with high modulation. Which in theory gives the best dose distribution in the body.

5.2 Discussing the correlations

Looking at the correlation between the parameters that influence the outcome of the metric and between the metric and passing rates, some moderate and weak correlations were observed. The monitor units have a moderate positive linear correlation with the complexity metric both from full irradiation plans as for the individual arcs inside those plans. These results are the same for both the Clinac™ and TrueBeam™ results. This means that the higher the complexity, the more monitor units are given to the patients. This was an expected result since monitor units per Gy is also a parameter that has been used to get an indicator of the complexity of a plan. [39]

For the relationship between the gamma passing rates, a negative linear relationship was expected, but this has only partially come true. For both the linacs, the absolute gamma passing rate (3%|3mm) gave a weak positive linear relationship with the complexity metric. While for the relative gamma passing rates, the correlations were weak and negative. This can be explained by the fact that in the relative gamma test the calibration of the EPID does not play a big role anymore, while it does for the absolute gamma passing rate where the calibrated units of the calculated and measured plans are compared to get the gamma passing rate of the arc. The diminished effect of the calibration of the EPID can also explain why there are more plans with low complexity failing in the absolute gamma test, while in the relative test, more plans with high complexity fail to reach the gamma passing rate threshold. The calibration can be good for high doses, but less for low doses resulting in more deviating results with low dose plans and thus for low complexity plans as these two have a linear correlation. Sometimes a verification measurement is cut short, which also results in significant differences in absolute dose measurements. Therefore it's better to analyse the relative gamma passing rates. [39]

This correlation for the absolute gamma test was twice as strong for the arcs on the TrueBeam™ compared to those on the Clinac™. This can be explained by the high overall passing rates observed for the TrueBeam™ plans with more than 75% of the plans scoring having a passing rate above 99% and 88.4% scoring above the passing rate threshold of 97%, while for the Clinac™ arcs only 40.1% score above this 97% threshold. A few low gamma passing rates for arcs with low complexity can then result in this higher positive correlation compared to that of the Clinac™ beams where the plans generally score lower on the gamma test. The better passing rates for TrueBeam™ could be explained by the fact

that the Clinac™ is an older model of linear accelerator than the TrueBeam™ which have newer technology resulting in better accuracy of the delivered irradiation plan compared to the Clinac™.

5.3 Receiver operating characteristics

The results of these ROC analyses follow the conclusions of the correlation tests. With the absolute gamma test with 3% dose difference and 3mm distance to agreement resulting in an area under the curve of 0.565. This is close to 0.5 and as explained in chapter 3.3.3, 0.5 is equal to doing a coinflip to decide the outcome based on the complexity metric, meaning that it has little predictive value. Looking at the ROC-curves of both relative gamma tests with 1mm DTA and 3% or 2% DD, they result in similar curves and AUC's. The passing rate for both was chosen differently, for 3% DD and 1mm DTA the gamma threshold was 95%, based on the 95% used by Younge et. al. For the 2%DD and 1mm DTA test a gamma threshold of 90% was used to pass or fail a plan, higher percentages resulted in most plans failing the test because the passing criteria are so strict. This 90% was still within the limits of the guidelines by the NCS to evaluate gamma passing rates. The ROC-curve for the passed or failed criteria gave a result similar to that of the absolute gamma test. This could be expected because of the different type of data used in one plot, that makes it difficult to get a correlation between them.[2], [15]

With the low AUC values and the ROC-curves being close to a line straight through the origin, it's not possible to set any thresholds based on the data from both the absolute gamma test and the passed or failed criteria. For the relative gamma tests, this is possible because of the better AUC and the ROC-curve following the y-axis close at the start of the curve. Based on the data from this curve, a threshold of 0.19 mm^{-1} will only predict 1.52% of the plans with a complexity higher than this value wrongly as a plan that will fail QA based on the 95% passing rate for 3%DD and 1mm DTA while correctly predicting 22.1% of the plans that will fail the QA test for these parameters. If the threshold is reduced to 0.18 mm^{-1}, the TPR increases to 36.5% while the FPR also increases to 7.58%. The relative 2% DD and 1mm DTA test with a 90% gamma passing threshold gave similar percentages with thresholds of 0.1 mm^{-1} higher. For 0.20 mm^{-1} the TPR is 20.2% with a FPR of 0.86%, while a 0.19 mm^{-1} complexity threshold results in a TPR of 33.6% and a FPR of 6.38%.

The ROC-curve of the absolute gamma test for the TrueBeam™ arcs has an AUC of 0.193 indicating that it predicts the opposite from the hypothesis of what is a determined with the parameter. For this research that would mean that it predicts low complexity plans as plans that fail the pre-treatment QA instead of plans with a high complexity like the relative gamma test predicted for the Clinac™-plans. This corresponds to the linear positive relationship as it was described in chapter 5.2. But the same remarks and concerns apply here, and a ROC-curve of relative gamma tests is needed to really draw conclusions from this.

Figure 22 shows a recreation of the graphs in figure 7 with the red line representing the True positives and the green line representing true negatives. This graph is made with the results of the relative gamma test with the 3% DD and 1mm DTA criteria since this had the best AUC and thus the best predictability out of all the different tests. As it was described in chapter 3.3.3 it is best that these graphs overlap as little as possible to make the best possible distinction between both groups. There is a significant overlap from the start of the graph until around the upper limits for complexity of 0.19mm^{-1} to 0.20mm^{-1} where the true negatives are almost zero, but the true positives are not. This is the point that the threshold can start differentiating best between both groups.

Because of the overlap at the start, it is not possible to set a threshold for which pre-treatment QA verifications would not be necessary anymore because of the low complexity. The metric can not differentiate enough between failing and passing tests at that point. But it is possible to set thresholds for high complexity values where it is possible to predict that a plan will likely fail the pre-treatment QA test. This can help in daily practice, because the plans with a complexity higher than this threshold can be re-planned or altered to have lower complexity scores, resulting in a lower chance that the plan

will fail. This can help to reduce the total time spend on a patient because a plan that does not pas the pre-treatment verification has to go through planning, verification on the linac and then the analysis in the portal dosimetry application all over again.

5.4 Average complexity scores

From the averages it is possible to determine the categories that have to be made when doing further research and when an institution wants to calculate thresholds to use in daily practice. The averages and the associated standard deviations also give a good indication of how complex an arc or beam is compared to previously planned arcs and beams.

A first observation can be made when looking at the averages for the arcs used to treat tumours in the prostate and rectum. The complexity of the arcs planned for the TrueBeam™ linac are higher than the averages for the Clinac™. An explanation for this can be found in the MLC composition of both types of linac. The Clinac™ has MLC banks with leaves of 5mm in the centre and 10mm on the outside, while the MLC mounted on the TrueBeam™ has leaves of 2.5mm in the centre and 5mm on the outside of the bank. Because of the narrower leaves these can move more precise around certain volumes, but they can also form even smaller apertures than the leaves of minimum 5mm in the MLC of the Clinacs™. For a tumour of the same size in the middle of the MLC block, the MLC of the TrueBeam™ will use double the amount of leaves and thus have more possible open sides. All these factors can contribute to the higher average complexities for TrueBeam™ plans. The same difference between TrueBeam™ and Clinac™ arcs can be seen in the four other treatment sites where both are used to treat that specific treatment site.

A second observation from table 5 in chapter 4.4 is the difference in complexity for different treatment techniques with intensity modulated plans. This manifests itself clearly in the averages for the TrueBeam™ arcs and beams for irradiations in the brain. Single IMRT beams have the lowest average complexity, followed by the standard VMAT arcs and the SRS arcs with the highest complexity. This can be explained by how each of these techniques work. The IMRT beams are at a static gantry angle when they are irradiated onto a patient, the position of organs at risk and healthy tissue is therefore constant relative to the position f the beam, requiring less repositioning of the MLC leaves. Compared to MLC or SRS arcs that usually contain 178 control points over an arc at different angles. The significant difference between VMAT and SRS arcs follows from the positive linear relationship between the amount of monitor units and the complexity metric. SRS is a treatment technique where a high amount of monitor units is delivered in one session, so the amount of monitor units of one arc is also high compared to VMAT arcs for normal irradiations. A second aspect that impacts the high complexity for SRS is that it uses small, very precise beams to treat in the brain. And small apertures result in higher complexity of the used arcs. Similar increases can be seen for SBRT for Lung irradiations, but the amount of MU’s in one session is lower than for SRS.

Another possible influence observed in the averages in table 5 is the difference in complexity for arcs used in the first part of the breast treatment and the subsequent boost plans to give extra dose to certain (smaller) parts in the breast region. Two differences can be observed between the arcs for the original therapy and the arcs used to deliver the boost. First is the difference in average MU's used, with 499.77 MU average for a boost arc, this is almost five times higher than the average amount of MU's for the original irradiations. This is because the entire dose is given by the arc in a boost plan, while during the standard irradiation a part of the dose is given by static fields. The amount of MU/Gy is important, because the dose that is given is significantly lower for a boost plan. Another difference is the amount of control points used in the arcs. The VMAT arcs for the original therapy mostly consist of 66 control points, while the boost arcs consist of 114 control points. To see if the amount of control points in the arc makes a difference, the arcs used to treat ORL tumours can be investigated. There are 3 major groups in these arcs: 98, 114 and 178 control points. All have a similar average complexity with only a difference of $0.002mm^{-1}$ between those averages. So, this does not affect the outcome of the metric.

5.5 Implementation and further research

To implement this metric in the daily workflow when planning intensity modulated treatment plans, the average of the metrics can be implemented the fastest. This can be used as reference to plan new irradiations. By aiming to achieve a complexity closer to the average if the complexity of an arc is high, a part of all the failed pre-treatment QA verifications can be avoided. The average should also go down if this is done over a period of time and new averages are calculated. Therefore, it can be interesting to review and update the averages periodically. The complexity thresholds of 0.18-0.19 mm^{-1} are good values to aim at to get below when planning VMAT arcs for the Clinac™. To achieve this for the other categories in table 5, ROC-curves have to be made for those as well. With a complexity that is lower on average, it is also possible that the groups of true positives and negatives will be separated more, resulting in better defined thresholds.

Important to consider is that the values that are determined are specific to treatment centres. Because the outcome of the calculations and curves depends on the type of linac and the MLC that is mounted on the head of the linac. The type of EPID detector also influences the outcome, because different EPID's have different specifications and perform differently, which results in other outcomes for the gamma score. Different software to calculate the portal dose image will also have an influence, since different programs can use different algorithms. When a new program is used to calculate the passing rates for the pre-treatment verifications and the intent is to use the complexity metric, it's important to keep track of the average complexity, then evolutions in the metric can be observed and compared to the data from the previous program. It can also be interesting to redo pre-treatment tests that were tested with the old program or equipment again with the new program or equipment. This way a possible relationship between the old and new acquired values can be found so the averages and thresholds that were previously determined can still serve as a reference when the new program or equipment is implemented.

The scripts that were made and used were only able to read data from servers, this data could be used or saved on the users end, like the complexity calculations or the readout of gamma passing rates from the portal dosimetry system. This last process is currently very time consuming, since it only reads out the gamma passing rate that was used to pass or fail the arc and therefore saved in the system. This is a combination of absolute gamma passing rates, relative passing rates with different normalization techniques and sometimes slightly different dose difference and distance to agreement criteria. To get other values, the patient file must be opened manually and the percentages have to be noted in a program like Excel. At the LOC, a new sort of scripting is being tested out that could help with this. This would allow to send information to the server, possibly allowing this process to be automated. That would help to analyse all the different treatment sites in a shorter period of time, analyse more patients and redo the calculations periodically without taking a lot of time.

The combination of different metrics can be in interesting route to do more research. Where the result of multiple complexity scores is used to determine if a plan is likely to pass or fail pre-treatment QA. Because as it was shown in chapter 2.4 in the literature study, different complexity metrics research different aspects that can contribute to the complexity of a plan and a possible failed pre-treatment QA test. The combination of different formulas also might be an option to try and get better results. The edge metric that was used in this research could perhaps be combined with the edge area metric by Götstedt et al. as described in paragraph 2.4.6 in this paper. While the edge metric only uses the length of the open leaf side, the edge area metric uses an area of 5 mm around the open MLC edges (sides and top of the MLC) encompasses the penumbra. In the research by Gödstedt et al. this method had better Pearson's R-values for gamma pass rates of EPID measurements, for passing rates determined with EBT3 film dosimetry, they had the same correlation for 3% DD, for 5% DD the Edge Metric scored better. Combined this would result in y_i in formula 14 being replaced by an area of 5mm to both sides of the open leaf side (without the top). The denominator A_i would be expanded with the area of 5mm under the MLC's.

6 Conclusion

In this research two scripts were successfully developed, one for use as a plug-in in Eclipse in daily practice when planning VMAT RapidArc™ plans and one to be used as a data gathering tool to analyse multiple irradiation plans at once.

It was shown that there is a positive linear correlation between the used edge metric that determines the complexity of a plan, single arc or beam and the amount of monitor units that was used to deliver the plan to the patient. A difference in results when using an absolute or relative method to generate gamma passing rates was ascertained. With the relative method delivering the best results for this research.

Thresholds were proposed at 0.18 and 0.19 mm^{-1} for which it is better to adjust or replan a treatment when the complexity is above these thresholds with different false positive and false negative rates. threshold. This is for the prostate arcs delivered with a Clinac™ at the Limburgs Oncologisch Centrum. For other treatment sites this was not possible due to a lack of data. For these, the average complexity calculated in this research can be used as a reference value to gain an insight into the complexity of the plan.

Differences in the outcome of the complexity metric were determined and described. Different treatment sites, treatment units and treatment methods must be distinguished from each other when doing research. Averages and eventually thresholds must be determined for each separate group to get the best results.

7 Bibliography

[1] K. C. Younge, M. M. Matuszak, J. M. Moran, D. L. McShan, B. A. Fraass, en D. A. Roberts, 'Penalization of aperture complexity in inversely planned volumetric modulated arc therapy', *Med. Phys.*, vol. 39, nr. 11, pp. 7160–7170, 2012, doi: 10.1118/1.4762566.

[2] K. C. Younge, D. Roberts, L. A. Janes, C. Anderson, J. M. Moran, en M. M. Matuszak, 'Predicting deliverability of volumetric-modulated arc therapy (VMAT) plans using aperture complexity analysis', *J. Appl. Clin. Med. Phys.*, vol. 17, nr. 4, pp. 124–131, 2016, doi: 10.1120/jacmp.v17i4.6241.

[3] 'Cancer'. https://www.who.int/news-room/fact-sheets/detail/cancer (geraadpleegd dec. 18, 2019).

[4] J. M. Borras *e.a.*, 'How many new cancer patients in Europe will require radiotherapy by 2025? An ESTRO-HERO analysis', *Radiother. Oncol.*, vol. 119, nr. 1, pp. 5–11, apr. 2016, doi: 10.1016/j.radonc.2016.02.016.

[5] J. B. Little, 'Principal Cellular and Tissue Effects of Radiation', *Holl.-Frei Cancer Med. 6th Ed.*, 2003, Geraadpleegd: feb. 17, 2020. [Online]. Beschikbaar op: https://www.ncbi.nlm.nih.gov/books/NBK12344/.

[6] T. D. L. F. Herman *e.a.*, 'Dosimetric comparison between IMRT delivery modes: Step-and-shoot, sliding window, and volumetric modulated arc therapy — for whole pelvis radiation therapy of intermediate-to-high risk prostate adenocarcinoma', *J. Med. Phys. Assoc. Med. Phys. India*, vol. 38, nr. 4, pp. 165–172, 2013, doi: 10.4103/0971-6203.121193.

[7] K. Iqbal, M. Isa, S. A. Buzdar, K. A. Gifford, en Muhammad. Afzal, 'Treatment planning evaluation of sliding window and multiple static segments technique in intensity modulated radiotherapy', *Rep. Pract. Oncol. Radiother.*, vol. 18, nr. 2, pp. 101–106, dec. 2012, doi: 10.1016/j.rpor.2012.10.003.

[8] M. T. M. Davidson, S. J. Blake, D. L. Batchelar, P. Cheung, en K. Mah, 'Assessing the role of volumetric modulated arc therapy (VMAT) relative to IMRT and helical tomotherapy in the management of localized, locally advanced, and post-operative prostate cancer', *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 80, nr. 5, pp. 1550–1558, aug. 2011, doi: 10.1016/j.ijrobp.2010.10.024.

[9] E. J. Hall, 'Intensity-modulated radiation therapy, protons, and the risk of second cancers', *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 65, nr. 1, pp. 1–7, mei 2006, doi: 10.1016/j.ijrobp.2006.01.027.

[10] Chang-Ming Charlie Ma, 'Physics and Dosimetric Principles of SRS and SBRT', *Mathews J. Cancer Sci.*, vol. 4, nr. 2, p. 16, 2019, doi: https://doi.org/10.30654/MJCS.10022.

[11] C. B. Saw, S. Bao, en S. Li, 'A review on the technical and dosimetric aspects of stereotactic body radiation therapy (SBRT)', *J. Radiat. Oncol.*, vol. 1, nr. 4, pp. 317–322, dec. 2012, doi: 10.1007/s13566-012-0025-z.

[12] J. M. Brown, D. J. Carlson, en D. J. Brenner, 'The Tumor Radiobiology of SRS and SBRT: Are More than the 5 R's Involved?', *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 88, nr. 2, pp. 254–262, feb. 2014, doi: 10.1016/j.ijrobp.2013.07.022.

[13] U. Schneider, E. Pedroni, en A. Lomax, 'The calibration of CT Hounsfield units for radiotherapy treatment planning', *Phys. Med. Biol.*, vol. 41, nr. 1, pp. 111–124, jan. 1996, doi: 10.1088/0031-9155/41/1/009.

[14] M. F. Chan, J. Li, K. Schupak, en C. Burman, 'Using a Novel Dose QA Tool to Quantify the Impact of Systematic Errors Otherwise Undetected by Conventional QA Methods: Clinical Head and Neck Case Studies', *Technol. Cancer Res. Treat.*, vol. 13, nr. 1, pp. 57–67, feb. 2014, doi: 10.7785/tcrt.2012.500353.

[15] A. Mans *e.a.*, 'The NCS code of practice for the quality assurance and control for volumetric modulated arc therapy', *Phys. Med. Biol.*, vol. 61, nr. 19, pp. 7221–7235, sep. 2016, doi: 10.1088/0031-9155/61/19/7221.

[16] A. Mans *e.a.*, 'Catching errors with in vivo EPID dosimetry', *Med. Phys.*, vol. 37, nr. 6Part2, pp. 2638–2644, 2010, doi: 10.1118/1.3397807.

[9] A. Vanbuel, "Implementatie van een top-down in-vivo QA-systeem", Ongepubliceerd eindwerk, UHasselt, Hasselt, 2019.

[18] D. I. Thwaites, 'QUALITY ASSURANCE OF EXTERNAL BEAM RADIOTHERAPY', p. 44.

[19] L. Masi, F. Casamassima, R. Doro, en P. Francescon, ‘Quality assurance of volumetric modulated arc therapy: Evaluation and comparison of different dosimetric systems’, *Med. Phys.*, vol. 38, nr. 2, pp. 612–621, 2011, doi: 10.1118/1.3533900.
[20] C. Collins *e.a.*, ‘An investigation of a novel reusable radiochromic sheet for 2D dose measurement’, *Med. Phys.*, vol. 46, nr. 12, pp. 5758–5769, 2019, doi: 10.1002/mp.13798.
[21] C. Wouter, V. Dirk, L. Paul, en D. Tom, ‘A reusable OSL-film for 2D radiotherapy dosimetry’, *Phys. Med. Biol.*, vol. 62, nr. 21, pp. 8441–8454, okt. 2017, doi: 10.1088/1361-6560/aa8de6.
[22] P. Casolaro *e.a.*, ‘Real-time dosimetry with radiochromic films’, *Sci. Rep.*, vol. 9, nr. 1, pp. 1–11, mrt. 2019, doi: 10.1038/s41598-019-41705-0.
[23] A. Grządziel, B. Smolińska, R. Rutkowski, en K. Ślosarek, ‘EPID dosimetry – configuration and pre-treatment IMRT verification’, *Rep. Pract. Oncol. Radiother.*, vol. 12, nr. 6, pp. 307–312, nov. 2007, doi: 10.1016/S1507-1367(10)60069-7.
[24] B. M. C. McCurdy en P. B. Greer, ‘Dosimetric properties of an amorphous-silicon EPID used in continuous acquisition mode for application to dynamic and arc IMRT’, *Med. Phys.*, vol. 36, nr. 7, pp. 3028–3039, 2009, doi: 10.1118/1.3148822.
[25] B. M. C. McCurdy, ‘Dosimetry in radiotherapy using a-Si EPIDs: Systems, methods, and applications focusing on 3D patient dose estimation’, *J. Phys. Conf. Ser.*, vol. 444, p. 012002, jun. 2013, doi: 10.1088/1742-6596/444/1/012002.
[26] J. W. Wong, ‘Electronic Portal Imaging Devices (EPID)’, in *Encyclopedia of Radiation Oncology*, L. W. Brady en T. E. Yaeger, Red. Berlin, Heidelberg: Springer, 2013, pp. 207–213.
[27] D. A. Low, W. B. Harms, S. Mutic, en J. A. Purdy, ‘A technique for the quantitative evaluation of dose distributions’, *Med. Phys.*, vol. 25, nr. 5, pp. 656–661, 1998, doi: 10.1118/1.598248.
[28] J. Hrbacek, ‘Routine verification of RapidArc® plans using EpiqaTM’, p. 11, 2010.
[29] M. Hussein, C. H. Clark, en A. Nisbet, ‘Challenges in calculation of the gamma index in radiotherapy – Towards good practice’, *Phys. Med.*, vol. 36, pp. 1–11, apr. 2017, doi: 10.1016/j.ejmp.2017.03.001.
[30] Netherlands Commission on Radiation Dosimetry, ‘Code of Practice for the Quality Assurance and Control for Intensity Modulated Radiotherapy’. jun. 2013.
[31] M. Hussein, P. Rowshanfarzad, M. A. Ebert, A. Nisbet, en C. H. Clark, ‘A comparison of the gamma index analysis in various commercial IMRT/VMAT QA systems’, *Radiother. Oncol.*, vol. 109, nr. 3, pp. 370–376, dec. 2013, doi: 10.1016/j.radonc.2013.08.048.
[32] F. Abolaban, S. Zaman, J. Cashmore, A. Nisbet, en C. H. Clark, ‘Changes in Patterns of Intensity-modulated Radiotherapy Verification and Quality Assurance in the UK’, *Clin. Oncol.*, vol. 28, nr. 8, pp. e28–e34, aug. 2016, doi: 10.1016/j.clon.2016.01.013.
[33] S. B. Crowe *e.a.*, ‘Treatment plan complexity metrics for predicting IMRT pre-treatment quality assurance results’, *Australas. Phys. Eng. Sci. Med.*, vol. 37, nr. 3, pp. 475–482, sep. 2014, doi: 10.1007/s13246-014-0274-9.
[34] M. Mathot en D. Dechambre, ‘6 VMAT complexity metrics can reduce patient QA workload’, *Phys. Med.*, vol. 56, pp. 3–4, dec. 2018, doi: 10.1016/j.ejmp.2018.09.019.
[35] A. L. McNiven, M. B. Sharpe, en T. G. Purdie, ‘A new metric for assessing IMRT modulation complexity and plan deliverability’, *Med. Phys.*, vol. 37, nr. 2, pp. 505–515, 2010, doi: 10.1118/1.3276775.
[36] S. Webb, ‘Use of a quantitative index of beam modulation to characterize dose conformality: illustration by a comparison of full beamlet IMRT, few-segment IMRT (fsIMRT) and conformal unmodulated radiotherapy’, *Phys. Med. Biol.*, vol. 48, nr. 14, pp. 2051–2062, jul. 2003, doi: 10.1088/0031-9155/48/14/301.
[37] G. Nicolini, A. Fogliata, en L. Cozzi, ‘IMRT with the sliding window: Comparison of the static and dynamic methods. Dosimetric and spectral analysis’, *Radiother. Oncol.*, vol. 75, nr. 1, pp. 112–119, apr. 2005, doi: 10.1016/j.radonc.2005.03.009.
[38] J. Llacer, T. D. Solberg, en C. Promberger, ‘Comparative behaviour of the Dynamically Penalized Likelihood algorithm in inverse radiation therapy planning’, *Phys. Med. Biol.*, vol. 46, nr. 10, pp. 2637–2663, sep. 2001, doi: 10.1088/0031-9155/46/10/309.
[39] J. Götstedt, A. K. Hauer, en A. Bäck, ‘Development and evaluation of aperture-based complexity metrics using film and EPID measurements of static MLC openings’, *Med. Phys.*, vol. 42, nr. 7, pp. 3911–3921, 2015, doi: 10.1118/1.4921733.

[40] I. J. Das, G. X. Ding, en A. Ahnesjö, 'Small fields: Nonequilibrium radiation dosimetry', *Med. Phys.*, vol. 35, nr. 1, pp. 206–215, 2008, doi: 10.1118/1.2815356.
[41] J. M. Park, S.-Y. Park, en H. Kim, 'Modulation index for VMAT considering both mechanical and dose calculation uncertainties', *Phys. Med. Biol.*, vol. 60, nr. 18, pp. 7101–7125, aug. 2015, doi: 10.1088/0031-9155/60/18/7101.
[42] L. Masi, R. Doro, V. Favuzza, S. Cipressi, en L. Livi, 'Impact of plan parameters on the dosimetric accuracy of volumetric modulated arc therapy', *Med. Phys.*, vol. 40, nr. 7, p. 071718, 2013, doi: 10.1118/1.4810969.
[43] 'Regression and Correlation'. https://internal.ncl.ac.uk/ask/numeracy-maths-statistics/statistics/regression-and-correlation/strength-of-correlation.html (geraadpleegd mei 02, 2020).
[44] S. Narkhede, 'Understanding AUC - ROC Curve', *Medium*, mei 26, 2019. https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5 (geraadpleegd mei 02, 2020).