University of Stavanger

Faculty of Science and Technology

BACHELOR`S THESIS

Study program/Specialization: Biological Chemistry	Spring semester, 2021 Open / Restricted access
Writer: Alina John	Alina John (Writer's signature)
Faculty supervisor: Hanne Røland Hagland	
Thesis title: Assessing the usage of predictive modeling to anticipate the response of cancer cells to metformin treatment	
Credits (ECTS): 20	
Keywords: Metformin treatment, Cancer metabolism, Viability, HCT116, Machine learning algorithms, Support Vector Machines	Pages: 30 + enclosure: 2 Stavanger, 15.05/2021 Date/year

Acknowledgments

First, I would like to express my utmost gratitude to my supervisor Hanne Røland Hagland for the opportunity to take part in this exciting and enlightening research project. I highly appreciate the guidance, support and feedback I have received.

The team at CORE has been exceptional and helped me acquire many new skills and techniques. I am very grateful to Julie Nikolaisen for teaching me various laboratory techniques and for her continued patience and assistance. I am also grateful to Abdelnour Alhourani for teaching me valuable skills in python and machine learning.

I had great pleasure working with Elisabeth Lassa and Hedvig Svensson on this project and am very appreciative of their cooperation and encouragement. Lastly, I would like to acknowledge the master's and PhD students at CORE for creating a warm and friendly atmosphere. I am especially thankful for their assistance and encouraging words.

Abstract

As artificial intelligence is steadily rising, the application of machine learning to predict drug responses of cancer cells can be of great value in cancer therapy. Although the drug metformin was originally developed to treat diabetes, its apparent effect on various cancer cells has been widely studied. This study primarily investigates the usage of machine learning algorithms to predict the response of cancer cells to metformin.

A database was constructed from published articles regarding the viability and metabolic changes in cancer cells after exposure to metformin. By applying linear regression to the database, a weak negative linear correlation of – 0.21 was observed between viability and metformin concentration.

Furthermore, various support vector machine algorithms were applied to find models that could make accurate predictions for viability according to the specific variables. From the varying levels of accuracy of the different SVM models, it was evident that the best-suited parameters and kernel functions must be selected to construct accurate models with high performance.

In this study, the colorectal cancer cell line, HCT116, was also exposed to metformin for 24 and 48-hour treatments to directly examine the drug's influence on cancer cells. To measure viability, alamarBlue assay and CCK-8 assays were conducted on cells treated with various concentrations of metformin. Here, it was apparent that exposure to metformin at higher concentrations for a longer period led to the greatest reduction in cancer cell viability and metabolic activity, displayed by the decreasing trend.

Overall, the usage of machine learning algorithms demonstrated its potential to make highly accurate models that could predict cancer cell's response to metformin treatment. This could further contribute to establishing new cancer treatments.

TABLE OF CONTENTS

1 Introduction

1.1 Cancer & Cancer Metabolism

While our body is constantly renewing cells by cellular division in a systematic manner, the cells that proliferate abnormally and uncontrollably can lead to a collection of diseases, commonly known as cancer. Although there have been many advancements in cancer therapy throughout the years, it remains a global health concern. Therefore, researchers are constantly seeking new approaches for its treatment.

One of the characteristics of cancer cells that distinguish them from normal cells is a shift in their metabolism. This is because cancer cells have the ability to undergo metabolic reprogramming to support their high energy demand (Dong and Neuzil 2019). Furthermore, this reprogramming supports the synthesis of essential biosynthetic precursors and the regulation of redox balance (DeBerardinis and Chandel 2016, p. 12). Thereby, these altered characteristics provide the resources and conditions necessary to sustain rapid cell growth and proliferation.

1.1.1 Glucose metabolism

The glycolytic activity is especially high in cancer cells and is a necessity for their continued growth and survival (Yu et al. 2017, p. 3430). This activity is primarily characterized by an altered metabolic trait that is frequently observed in these cells, which is their ability to perform aerobic glycolysis. While normal cells only metabolize glucose to lactate under anaerobic conditions, cancer cells metabolize glucose to lactate even under aerobic conditions (Kalyanaraman 2017, p. 835). This altered characteristic is known as the Warburg effect. A second altered metabolic trait exhibited by cancer cells is their ability to take up high levels of glucose. This is due to the overexpression of certain glucose transporters (GLUTs), primarily GLUT1 and GLUT3 (Kalyanaraman 2017, p. 835). The overexpression of GLUTs is often induced by the activation of oncogenes such as *C-MYC*, *RAS* and *SRC* (Adekola et al. 2012).

Although glycolysis only yields 2 ATP molecules per molecule of glucose, their ability to perform aerobic glycolysis at a high rate provides a continuous generation of ATP and key intermediates (Alam et al. 2016). The glycolytic intermediates drive the pentose phosphate pathway, which leads to the generation of ribose-5-phosphate and NADPH (Anderson et al.

2018b, p. 998). These intermediates are necessary to synthesize lipids and nucleic acids, which further supports cancer cell proliferation. (Ganapathy-Kanniappan and Geschwind 2013, p. 3). Overall, the performance of aerobic glycolysis at a high rate sustains the generation of ATP while simultaneously providing important intermediates that are necessary for successful cell proliferation and growth.

1.1.2 Mitochondrial metabolism

As the mitochondria exhibit a variety of functions that are both beneficial and necessary for the growth and survival of cancer cells, it is a fundamental organelle in cancer metabolism and tumorigenesis. In addition to being the major source of ATP through oxidative phosphorylation, the mitochondria generate biosynthetic intermediates that are needed for the growth of cancer cells (Liu and Shi 2020). It is explicitly the tricarboxylic acid (TCA) cycle that generates these in the mitochondrial matrix. These intermediates are essential elements required to synthesize macromolecules and to produce electron acceptors and energy (Anderson et al. 2018, p. 217). As depicted in Figure 1, the intermediates from the TCA cycle are used to synthesize cholesterol, fatty acids, amino acids and nucleotides, which are necessary elements to sustain cell growth and proliferation of cancer cells.

The mitochondria also produce reactive oxygen species (ROS) which can activate potential tumorigenic pathways (Porporato et al. 2018, p. 266). These species are also involved in inducing oxidative damage and modifying gene expression, which are necessary for the progression of cancer (Yang et al. 2016). However, excessive accumulation of ROS will be fatal to cancer cells. To prevent this, copious amounts of NADPH are generated in the mitochondria as this will limit the further accumulation of ROS (Weinberg and Chandel 2015). The amount of ROS is also kept at stable levels by the expression of antioxidants (DeBerardinis and Chandel 2016, p. 8). Accordingly, cancer cells require a balance in ROS to prevent potential harmful effects.

The electron transport chain (ETC) in the inner mitochondrial membrane also contributes to tumorigenesis. Namely Complex I of the ETC. As the main role of Complex I is to oxidize NADH to NAD^+, it elevates and maintains the NAD^+/NADH ratio in the mitochondrial matrix

(Santidrian et al. 2013, p. 1068). Additionally, Complex I supports cancer cell proliferation by supplying electron acceptors and by regenerating cofactors (Urra et al. 2017).

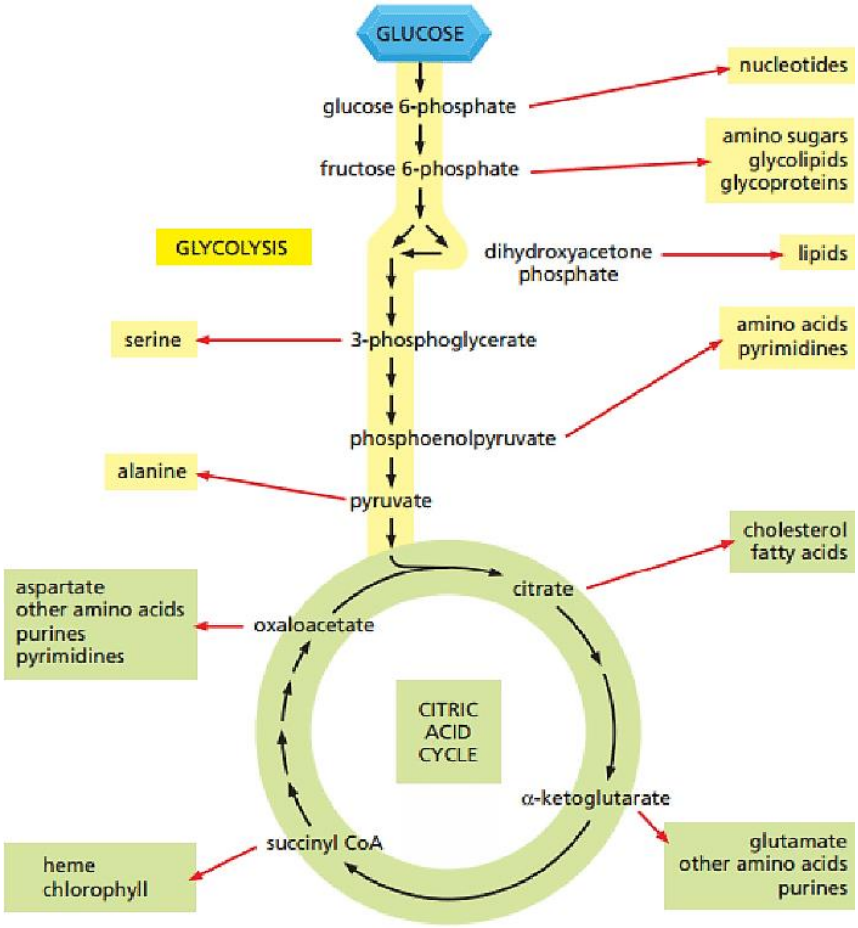

Figure 1: An overview of glycolysis and tricarboxylic acid, including the essential products produced by the glycolytic and TCA intermediates. (Nitzsche and Nishith Gupta 2017)

1.1.3 Targeting the mitochondria

Considering the various roles of mitochondria in cancer metabolism, targeting this organelle would be a clever tactic to limit cancer cell proliferation. Generally, such an approach would display inhibited cell respiration, limited proliferation and reduced ATP production (Kalyanaraman 2017, p. 840). By particularly targeting Complex I of the ETC, the NAD^+/NADH ratio would become disproportionate. This would lead to (1) reduction in electron acceptors, (2) reduced synthesis of aspartate, (3) limited purine and pyrimidine, (4) lack of materials needed to synthesize nucleic acids and macromolecules (Urra et al. 2017). These factors would thereby provide inadequate conditions for cell proliferation, resulting in its limitation.

1.2 Colorectal Cancer

1.2.1 What is colorectal cancer?

Colorectal cancer (CRC) is a form of cancer that occurs when growths called polyps arise in the colon or rectum, which has the potential to become cancerous over time (What Is Colorectal Cancer? | CDC 2021). It is the third most prevalent gastrointestinal cancer in the world (Liu et al. 2020, p. 3876). The development of CRC emerges gradually as genetic and epigenetic changes accumulate (Binefa et al. 2014). These changes appear in the epithelial cells of the colon, which eventually results in the formation of adenocarcinomas (Lao and Grady 2011).

1.2.2 Using cell lines as a model for cancer

A cell line of colorectal cancer, known as HCT116, will be used as a model for cancer disease in this study. The diversity of cancer makes it challenging to understand the disease. Therefore, the usage of models to represent cancer disease is very useful when conducting research. More specifically, using cancer cell lines can help evolve our knowledge regarding mechanisms of cancers, which can then be used to develop treatment methods. Scientists have debated if cancer cell lines are an appropriate and truthful model of the disease (Ferreira. D et al. 2013). However, it seems that cell lines can preserve genetic properties of the cancer of origin under proper conditions. (Mirabelli et al. 2019). Research has also suggested that cell lines from an early stage and lower grade of cancer represent more realistic models (Ferreira. D et al. 2013). These factors should be taken into consideration when deriving cell lines, to ensure that the cell line is a representative model for cancer disease.

1.3 Metformin

Although the drug metformin is mainly prescribed to treat type II diabetes mellitus, it has displayed the ability to influence certain cancer cells. Studies have suggested that metformin decreases the development of cancer in diabetes patients while also increasing the survival rate among cancer patients (Weinberg and Chandel 2015).

Metformin particularly influences the mitochondria of cancer cells and accumulates within the mitochondrial matrix due to its positive charge (Owen et al. 2000, A611). Here, it acts as a complex I inhibitor in the electron transport chain (ETC). Although the exact mechanism for this

inhibition is somewhat obscure, it has been proposed that the cysteine-39 loop in the ND3 subunit of complex I may be the binding site for metformin (Vial et al. 2019, p. 3).

The inhibition of complex I by metformin have numerous consequences on cancer cells. Firstly, it results in rising AMP and declining ATP. The increase in the AMP/ATP ratio leads to the activation of adenosine monophosphate-activated protein kinase (AMPK) which promotes catabolism and inhibition of fatty acid synthesis (Anderson et al. 2018b). As fatty acids are essential for the assembly of cellular membranes in rapidly proliferating cancer cells, this results in inhibited tumor growth (Vancura et al. 2018). Secondly, the oxygen consumption rate is decreased due to the inhibition of the ETC. Thirdly, the oxidation of NADH to NAD^+ is inhibited (Vancura et al. 2018). This leads to the accumulation of NADH in the matrix. Studies have also shown that the inhibition of complex I lead to reduced cell proliferation when glucose is present, and cell death when glucose is limited (Wheaton et al. 2014).

The decrease in ATP production triggered by metformin also exhibits other inhibitory effects on cancer cells. A kinase known as mTORC1 is frequently present in signaling pathways of cancer cells and is essential for the proliferation of cells (Vancura et al. 2018). However, the decrease in ATP production leads to diminished mTOR activity (Weinberg and Chandel 2015). This decline in mTORC1 results in decreased cell growth and proliferation. Also, increased AMP/ATP ratio inhibits pyruvate carboxylase, which further results in the inhibition of gluconeogenesis (Owen et al. 2000, p. 613).

1.4 Databases & Machine Learning

1.4.1 Machine Learning

To predict the response of cancer cells to metformin, machine learning (ML) algorithms that involve analyzing and learning data can be applied to generate mathematical models. ML algorithms are capable of recognizing significant patterns within given data and are therefore considered to be a form of artificial intelligence (Erickson et al. 2017, p. 505). ML is primarily divided into two categories: supervised learning and unsupervised learning (Sidey-Gibbons and Sidey-Gibbons 2019). In this study, the main focus will be supervised learning which requires both input data (features) and desired output data to train the algorithm (Goel 2018). To achieve successful training of the algorithm, the data should be separated into training data and testing

data. Both data will consist of a random selection of features and their outcomes (Sidey-Gibbons and Sidey-Gibbons 2019). Data enrichment and labeling are also necessary for the training data (CloudFactory). First, the algorithm is trained by understanding the relationship between selected features and their outcomes. By applying this trained algorithm to the testing data, their outcomes can be predicted based on the relationships observed in the training data. Thereby, the accuracy of the model can be determined.

There have been applied various supervised machine learning models to generate accurate drug response predictions (Huang et al. 2017). These include support vector machines (SVM), linear regression and decision trees. Support vector machine is a model that can generate predictions of both linear and non-linear data (Uddin et al. 2019). In this model a variety of kernel functions can be applied, including linear kernel, Gaussian radial basis (RBF) kernel and polynomial kernel (Savas and Dovis 2019). It is essential to choose the best-suited kernel function according to an individual database to ensure optimal performance. Frequently, RBF kernels are selected as they are less likely to be affected by noise and function well for both smaller and larger databases (Sun 2019). Linear regression is a model that assumes that there is a linear relationship between the input variables (x) and the output variable (y) (Olawale 2020). Thus, this model can be used to predict the relationship between various cancer metabolism parameters and metformin exposure. Decision trees are another frequently used method of supervised machine learning. This model is constructed as a tree and involves decision-making. The tree begins at the root node, which branches into internal nodes, representing the possible choices, until it reaches the leaf nodes that contain the final outcomes (Song Yan-yan and LU 2015, p. 131).

As evidence has suggested that cancer pathways are not necessarily defined by the origin of the tumor, applying machine learning models across various cancer types generates more accurate predictions (Huang et al. 2017)). Accordingly, one should not explicitly use data from specific cancers when building models to predict the drug response of a particular form of cancer.

1.4.2 Databases

These various supervised machine learning models can be applied to specifically predict the response of cancer cells to metformin treatment. First, data collection involving the responses of

various cancers to metformin must be collected to train the algorithm. As stated earlier, the collected data should be from a wide variety of cancer forms, to ensure accurate predictions.

In order to apply the machine learning algorithms, the database must be imported into Pandas DataFrame which uses the programming language Python. By importing the collected data, the databases get organized into a format that is appropriate for analysis. This format involves a two-dimensional structure consisting of labeled rows and columns, aligned by arithmetic operations (pandas.DataFrame 2021).

1.5 Seahorse Analyzer

As the drug metformin seems to influence the mitochondria in cancer cells, a Seahorse XF Analyzer can be used to measure changes in their mitochondrial activity after metformin exposure. Seahorse XF Analyzer is an instrument that can directly measure fundamental mitochondrial parameters, including oxygen consumption rate (OCR), extracellular acidification rate (ECAR) and ATP production (Agilent Technologies, Inc., p. 5).

In this instrument, OCR and ECAR are measured while successively adding the electron transport chain (ETC) inhibitors oligomycin, FCCP, Rotenone, and antimycin A (Tan et al. 2015, p. 240). While oligomycin inhibits ATP synthase (complex V), FCCP disturbs the mitochondrial membrane potential, and Rotenone and antimycin A inhibits complex I and III, respectively (Kalyanaraman et al. 2018, p. 318). By adding these ETC inhibitors one can measure the mitochondrial parameters of basal respiration, ATP-linked respiration, proton leak, maximal respiratory capacity and non-mitochondrial oxygen consumption.

The changes in the oxygen consumption rate after the addition of the ETC inhibitors can be observed in figure 2. Here, the inhibition of ATP synthase by oligomycin and the inhibition of Complex I and II by Rotenone and antimycin A leads to a rapid decline in OCR. Meanwhile, the addition of FCCP causes a rapid increase in OCR.

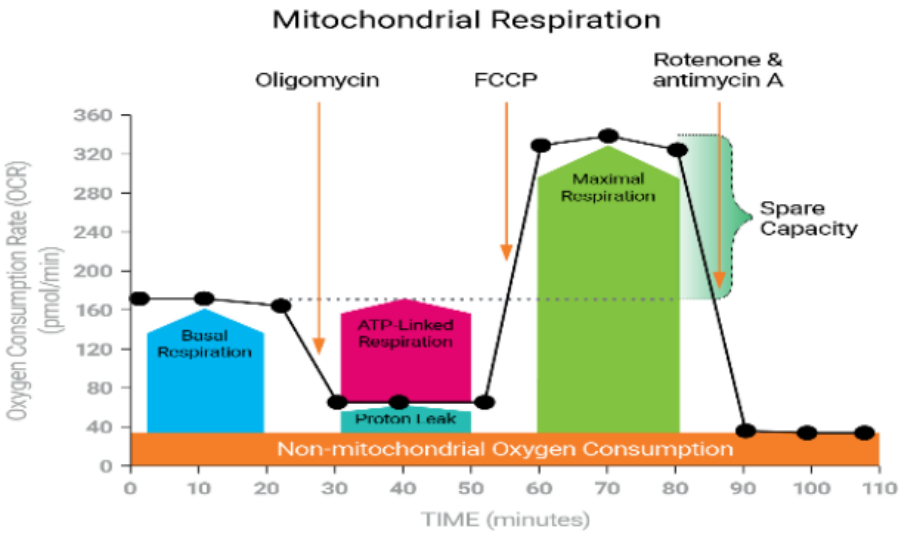

Figure 2: The change in oxygen consumption rate (OCR) after the successive addition of the ETC inhibitors: oligomycin, FCCP, Rotenone and antimycin A. These additions show the measurements for the mitochondrial parameters: basal respiration, ATP-linked respiration, proton leak, maximal respiratory capacity and non-mitochondrial oxygen consumption. Image: (Agilent Technologies, Inc., p. 5)

1.6 Aims

The primary aim of this study is to build a database using peer-reviewed articles involving the effect of metformin on the viability and mitochondrial activity of cancer cells. It is intended to further use this database to train machine learning algorithms to assess their ability to make predictions. The secondary aim is to examine the effects of metformin on cancer cells by conducting various viability tests on a colorectal cancer cell line treated with metformin.

2 Materials & Methods

2.1 Cell Line & Culture

2.1.1 Cell Line & Medium Preparation

The human colonic cancer cell line, HCT116, was obtained from the European Collection of Authenticated Cell Cultures (ECACC). This particular cell line was a strain of malignant cells isolated from a male with colonic carcinoma. (ECACC General Cell Collection: 91091005 HCT 116). The cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM) enriched with fetal bovine serum, penicillin: streptomycin solution, L-Glutamine, and glucose. The appropriate volumes of the various components are presented in Table 1.

Table 1: The components and volumes used to prepare a complete Dulbecco's Modified Eagle's Medium.

Components	**Volume**
Dulbecco's Modified Eagle's Medium without glucose, L-Glutamine and sodium pyruvate – Corning®	500 ml
Fetal bovine serum, heat-inactivated, South American origin – Biowest	50 ml
Penicillin : Streptomycin solution 6,0/10,0 g/L 100X – Biowest	5 ml
L-Glutamine, 200mM – Corning®	5 ml
Glucose, 2.5 M	2.22 ml (1g/L)

2.1.2 Cell Culture Maintenance

For cell culture maintenance in the T75 flask, the medium was replaced every two days and the HCT116 cell suspension was split 1:10 at 50% confluence. The components required for maintenance, complete DMEM, PBS and trypsin, were all warmed in a water bath at 37 °C for 5 minutes before use. The PBS solution was made by dissolving a 5 g Gibco® PBS Tablet in 500 mL of distilled water, followed by autoclaving.

When splitting the suspension 1:10, the present medium was aspirated, and 1 ml PBS was gently swished over the cells for washing. After aspirating the PBS, 1 ml of Trypsin EDTA 1X (0.25% Trypsin, 2.21 mM EDTA) from Corning® was added. The flask was then rocked to distribute the trypsin, followed by incubation at 37 °C for 1 minute until the cells detached from the flask wall. Thereafter, 4 ml complete DMEM was added and mixed with the cells by pipetting up and down. 4.5 ml of the cell suspension was aspirated, leaving 0.5 ml in the flask. 9.5 ml fresh medium (DMEM) was added and the flask was left to incubate at 37 °C.

When transferring the cell suspension to a new T75 flask, the steps above were repeated but rather than leaving 0.5 ml suspension in the old flask it was extracted and transferred to the new flask. The 9.5 ml fresh medium was then added to this new flask before incubation.

2.2 Muse® Count & Viability Reagent

2.2.1 Principle

The Muse Count and Viability Kit was used to determine cell count & viability. As the Muse Count & Viability Reagent contains two different DNA staining dyes, it distinguishes viable cells from non-viable cells. While the DNA-binding dye stains the nucleus of dead/dying cells that have lost membrane integrity, the membrane-permanent DNA staining dye stains all nucleated cells to distinguish cells from debris (Muse® Count & Viability Kit (200X) User's Guide 2020, p. 2)

2.2.2 Procedure

HCT116 cell suspension (25 µl) was added to a capless microcentrifuge tube along with Muse Count & Viability Reagent (225 µl). The sample was incubated for 5 minutes at room temperature to allow for cell staining to occur. After mixing the sample using a vortex mixer, the tube with stained cells was analyzed using the Guava Muse Cell Analyzer. The results from this procedure were used to make proper volume calculations for the alamarBlue Assay experiment.

2.3 Alamarblue® Cell Viability Assay Reagent

2.3.1 Principle

To find suitable seeding densities for the cell line HCT116, the alamarBlue viability test was conducted using the alamarBlue Cell Viability Reagent, which contains the REDOX indicator resazurin. This indicator displays fluorescence and colorimetric change from blue (oxidized form) to pink (reduced form), resulting from the metabolic activity caused by cellular growth (ThermoFisher, alamarBlue® Assay).

2.3.2 Procedure

The HCT116 cells were split using Trypsin in a T75 flask and the suspension (4.5 ml) was poured into a 15 ml tube. This tube was centrifuged for 5 minutes to separate the cell pellet. The remaining medium was extracted from the tube. A new calculated volume of medium was added using results from the Muse count viability test. Appropriate cell suspension volumes for 10 000,

15 000, 20 000, 25 000 and 30 000 viable cells/mL were found by further calculations. These were added to six Eppendorf tubes numbered 0, 10, 15, 20, 25 and 30.

Thereafter, cell suspensions of 100 µl were added to their assigned wells in a 96-well plate with increasing cell number downwards. PBS (110 µl) was added to all remaining wells except for the four corner wells. The cell plate was incubated at 37 °C for 48 hours. After incubation, alamarBlue (300 µl) was added into an Eppendorf tube and incubated at 37 °C for 5 minutes. 10 µl of this reagent was then added to the wells containing cell suspension and the content was mixed by gently tapping the plate. Resorufin, positive control (110 µl), was added to the four corner wells.

After 4 hours of incubation at 37 °C, the plate was analyzed using a SpectraMax Paradigm Multi-Mode Microplate Reader. Here, the fluorescence was measured with excitation wavelength at 540 nm and emission wavelength at 590 nm. The results were plot into excel to generate graphs.

2.4 Cell Counting Kit-8 (CCK-8)

The Cell Counting Kit-8 (CCK-8) was used to determine cell viability in HCT116 cells subjected to metformin treatment. This viability assay contains the highly water-soluble tetrazolium salt (WST-8), which can be reduced by dehydrogenases in cells to form the orange product formazan. From this, the number of viable cells can be determined by the intensity of the orange color, as they are proportional to one another (Dojindo, Cell Counting Kit-8 - Technical Manual).

2.5 Metformin Treatment

2.5.1 Preparation of Metformin

1.4 ml metformin stock solution of 250 mM was prepared in MQ water. Thereafter, 1 mM, 5 mM and 10 mM metformin solution was made in complete medium.

The following formulas were used for the calculations:

$$n = C \; x \; V \qquad (1)$$

$$m = n \; x \; M \qquad (2)$$

$$C1V1 = C2V2 \qquad (3)$$

Moles of metformin was calculated using formula (1):

$$n = 0.250\ x\ 0.0014\ L = 0.00035\ moles$$

Mass of metformin was calculated using formula (2):

$$m = 0.00035\ x\ 129.164 = 0.0452074\ grams$$

1 mM metformin using formula (3): (repeated for 5 mM and 10 mM)

$$250\ x\ V1 = 1\ x\ 1.4 \qquad V1 = \frac{1\ x\ 1.4}{250} = 0.0056\ ml = 5.6\ \mu l$$

$$1\ mM\ Metformin = 1394.4\ \mu l\ \text{medium} + 5.6\ \mu l\ \text{Metformin}$$

2.5.2 Procedure

The HCT116 cells were seeded at a density of 5.0×10^3 cells/well in two 96-well plates. This was determined as an appropriate seeding density by the previous AlamarBlue assay experiments. It was added 5000 cells/100µl to 30 wells in each plate. After allowing the cells to attach for 24 hours, the medium was removed from these wells.

Thereafter, fresh complete DMEM was added along with the various treatments to a total volume of 100µl/well. It was added six replicates of each treatment. This included six replicates of complete DMEM, complete DMEM with MQ water, 1 mM Metformin, 5 mM Metformin and 10 mM Metformin, as shown in Figure 3. To the four corner wells 110µl positive control (Resorufin) was added. 110µl PBS was added to all the remaining wells.

The first 96-well plate was treated for 24 hours. Thereafter, the suspensions in the 30 wells were subjected to two different viability assays, alamarBlue and CCK-8, after removing the treatments. AlamarBlue (10µl) and medium (100µl) were added to the top three replicates of each treatment, indicated by the blue color in Figure 3. CCK-8 (10µl) and medium (100µl) were added to the bottom three replicates of each treatment, indicated by the pink color in Figure 3. The second plate was treated for 48 hours before performing the same viability assays as explained above.

After performing the viability assays, the plates were incubated for 4 hours of incubation at 37 °C. Thereafter, they were analyzed in a SpectraMax Paradigm Multi-Mode Microplate Reader.

Here, the fluorescence of the cells treated with alamarBlue was measured with excitation wavelength at 540 nm and emission wavelength at 590 nm. The absorbance of the cells treated with CCK-8 was measured at 450 nm.

	1	2	3	4	5	6	7	8	9	10	11	12
A	CTRL	PBS	PBS	PBS	PBS	PBS	PBS	PBS	PBS	PBS	PBS	CTRL
B	PBS	DMEM	DMEM & MQ	1 mM MET	5 mM MET	10 mM MET	PBS	PBS	PBS	PBS	PBS	PBS
C	PBS	DMEM	DMEM & MQ	1 mM MET	5 mM MET	10 mM MET	PBS	PBS	PBS	PBS	PBS	PBS
D	PBS	DMEM	DMEM & MQ	1 mM MET	5 mM MET	10 mM MET	PBS	PBS	PBS	PBS	PBS	PBS
E	PBS	DMEM	DMEM & MQ	1 mM MET	5 mM MET	10 mM MET	PBS	PBS	PBS	PBS	PBS	PBS
F	PBS	DMEM	DMEM & MQ	1 mM MET	5 mM MET	10 mM MET	PBS	PBS	PBS	PBS	PBS	PBS
G	PBS	DMEM	DMEM & MQ	1 mM MET	5 mM MET	10 mM MET	PBS	PBS	PBS	PBS	PBS	PBS
H	CTRL	PBS	PBS	PBS	PBS	PBS	PBS	PBS	PBS	PBS	PBS	CTRL

AlamarBlue

CCK-8

Figure 3: Layout of the 96-well plate and the various treatments performed on HCT116 cells (5000 cells/well). The treatments were DMEM (control), DMEM & MQ water, 1 mM Metformin, 5 mM Metformin and 10 mM Metformin. Two different viability assays were conducted, AlamarBlue (blue) and CCK-8 (pink), on three replicates of each treatment.

2.6 Databases & Machine Learning

For data collection, an assortment of published peer-reviewed articles, exploring metformin as an anticancer agent, were studied. The articles mainly involved exposing metformin-treated cancer cells to a mitochondrial stress test using a Seahorse Analyzer. In addition, the viability of these cells was included. Relevant data were collected from the selected data variables and plotted into Excel. The selected data variables were Metformin concentration; Treatment time; Media glucose concentration; Viability; OCR basal; OCR oligo; OCR FCCP; OCR ROT/AA; ECAR.

To analyze the collected data, it was imported into Pandas DataFrame in Python using Google Colaboratory. Thereafter, various machine learning models were applied to assess their performance and accuracy. First, a linear regression model presenting the relationship between metformin concentration and viability was constructed using python and machine learning. This was to assess any existing trends that could be further used to make predictions regarding the cellular response to metformin. The correlation between each variable in the dataset was also found by creating a correlation matrix. To train support vector machine algorithms, 70 % of the

data was divided into training sets, while 30 % of the data was divided into testing sets. Thereafter, SVM with various kernels was applied and the accuracy of each model was determined.

3 Results

3.1 AlamarBlue Assay – Seeding Density

To find the appropriate seeding density of HCT116 for the metformin treatments, AlamarBlue assay was performed on cells of various seeding densities. It was observed that the fluorescence rapidly increased until they reached a seeding density of approximately 10 000 cells/well (Fig.4). Thereafter, the fluorescence remained quite steady with only slight rises at various points. As the graph particularly showed a rapid increase at 5000 cells/well, it was determined as an appropriate seeding density to use for the upcoming metformin treatments.

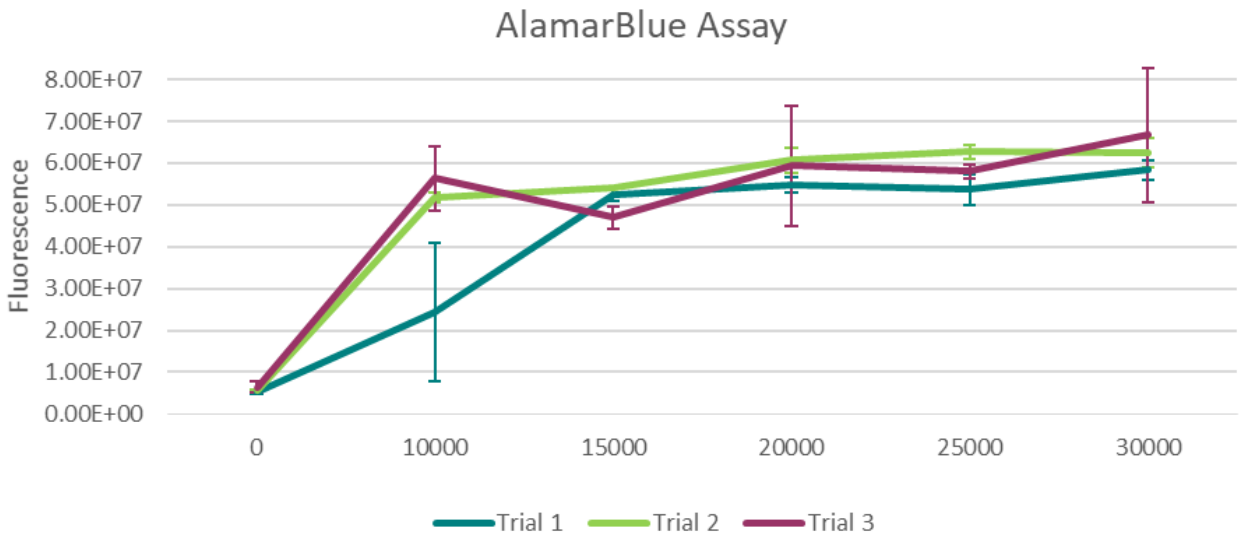

Figure 4: Measurement of fluorescence in HCT116 cells seeded at 0, 10000, 15000, 20000, 25000 and 30000 cells/well. The cell suspensions were incubated at 37°C for 48 h, before being incubated with AlamarBlue Viability Assay reagent for 4 h. Fluorescence was measured at wavelengths 540 nm (excitation) and 590 nm (emission) using a SpectraMax Paradigm Multi-Mode Microplate Reader. Three trials were performed.

3.2 Metformin Treatment of HCT116

The influence of metformin on the metabolic activity and viability of HCT116 cells was determined by performing AlamarBlue and CCK-8 assays to measure fluorescence and

absorbance, respectively. In the first trial, there was an evident decreasing trend in cell viability as metformin concentration increased, although some inconsistencies were observed (Fig.5).

After 24 hours of treatment using AlamarBlue assay, the viability displayed a slow and gradual decreasing trend (Fig.5A). The cells treated with 10 mM metformin diverged slightly from this trend as they displayed somewhat higher viability than those treated with 5 mM metformin. Meanwhile, after 48 hours of treatment, there was a more pronounced decrease in viability.

Viability measured after 24 hours using CCK-8 assay also displayed a slow and steady decrease (Fig.5B). There were some inconsistencies as the viability increased slightly at 5 mM metformin treatment before decreasing again at 10 mM metformin treatment, most likely due to technical errors. After 48 hours of treatment, there was a rapid decrease in viability of cells treated with 1 mM metformin compared to treatment with MQ water. However, this decrease stabilized as the metformin concentration was further increased.

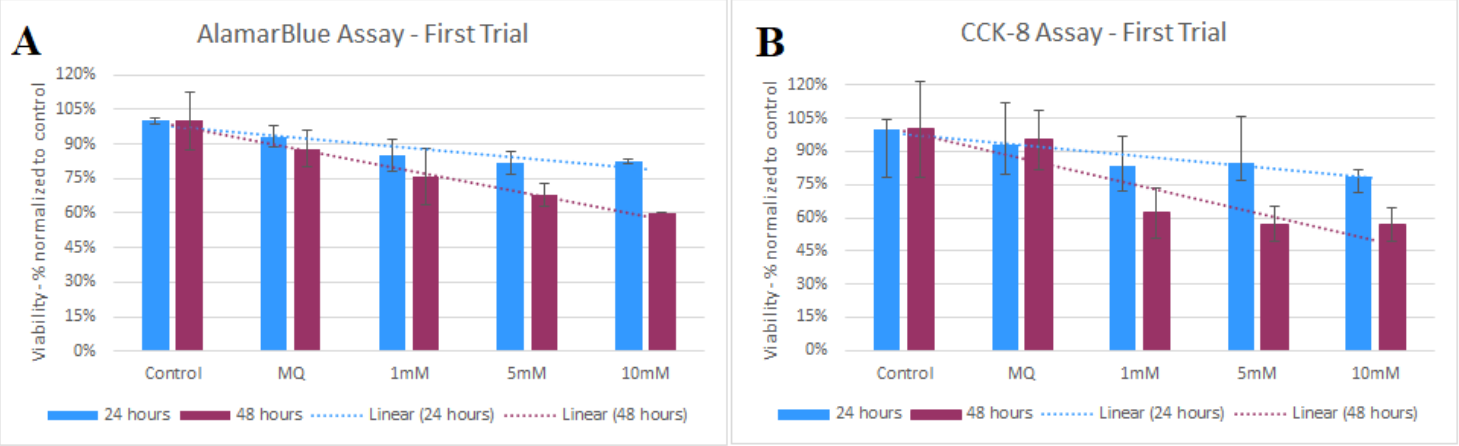

Figure 5: First trial of seeding HCT116 cells with MQ water, 1mM metformin, 5mM metformin and 10 mM metformin for 24 h (blue) and 48 h (purple). The cell viability was determined by performing AlamarBlue Assay (A) and CCK-8 Assay (B). All values were normalized to control. Relative standard deviation (%) is included.

In the second trial, the viability displayed somewhat irregular trends (Fig.6). Here, the measurements after 24 hours using AlamarBlue Assay showed how viability decreased in cells treated with 1 mM and 5 mM metformin while increasing after treatment with 10 mM (Fig.6A). There was also an increase in viability after MQ treatment compared to control. In the viability measured after 48 hours, the cells treated with MQ water and 1 mM MET exhibited higher

viability than the control. Thereafter, the viability decreased gradually at higher metformin concentrations.

The viability measured after 24 hours using CCK-8 Assay displayed a slowly decreasing trend, apart from cells treated with 10 mM metformin that exhibited slight elevations in viability (Fig.6B). In cells treated for 48 hours, the viability decreased at higher concentrations of metformin. However, it appears that cells treated with MQ water displayed a greater decrease in viability in comparison to those treated with 1 mM metformin.

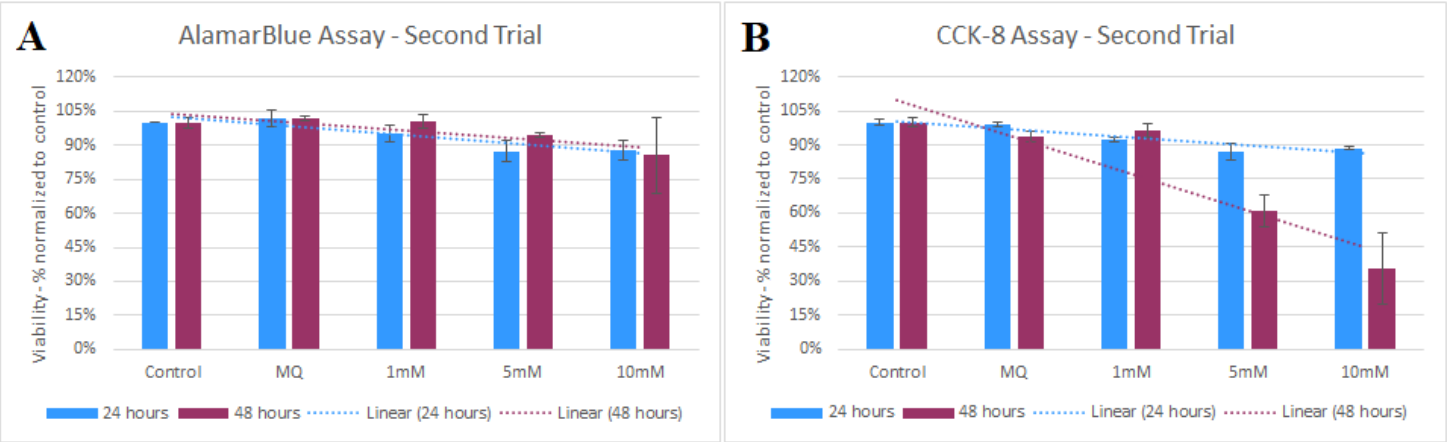

Figure 6: Second trial of seeding HCT116 cells with MQ water, 1mM metformin, 5mM metformin and 10 mM metformin for 24 h (blue) and 48 h (purple). The cell viability was determined by performing AlamarBlue Assay (A) and CCK-8 Assay (B). All values were normalized to control. Relative standard deviation (%) is included.

3.3 Databases – Machine Learning Models

Table 2: The variables in the database used to generate machine learning models, including descriptions of all its statistical functions.

	MetConcentration	TreatmentTime	MediaGlucose	Viability	OCRBasal	OCROligo	OCRFCCP	OCRROT/AA	ECAR
count	377.000000	313.000000	252.000000	189.000000	185.000000	144.000000	142.000000	140.000000	67.000000
mean	5.599574	39.504792	16.086429	88.470899	109.704865	84.911458	179.551761	122.957500	42.780597
std	14.664283	25.773514	10.962579	23.994358	133.130303	105.757327	244.966701	155.632014	47.912930
min	0.000000	0.250000	0.000000	11.000000	0.200000	0.000000	0.000000	0.000000	1.400000
25%	0.000000	24.000000	5.500000	77.000000	22.000000	25.000000	50.000000	31.500000	18.500000
50%	0.250000	24.000000	11.000000	95.000000	70.000000	41.500000	110.000000	90.000000	32.000000
75%	4.000000	48.000000	25.000000	100.000000	130.000000	100.000000	210.000000	164.000000	50.000000
max	75.000000	144.000000	100.000000	185.000000	750.000000	450.000000	1500.000000	1490.000000	350.000000

To achieve a better understanding regarding the effect of metformin concentrations on cell viability, their relationship was examined by a linear regression model developed using machine

learning (Fig.7). The model displayed a slight negative correlation between these variables. By developing a correlation matrix in Python, the correlation coefficients between each variable in the dataset were established (Fig.8). The correlation coefficient (r) between viability and metformin concentration was found to be -0.21. As expected, this matrix also displays the positive correlation between the OCR values from the mitochondrial stress test using Seahorse Analyzer.

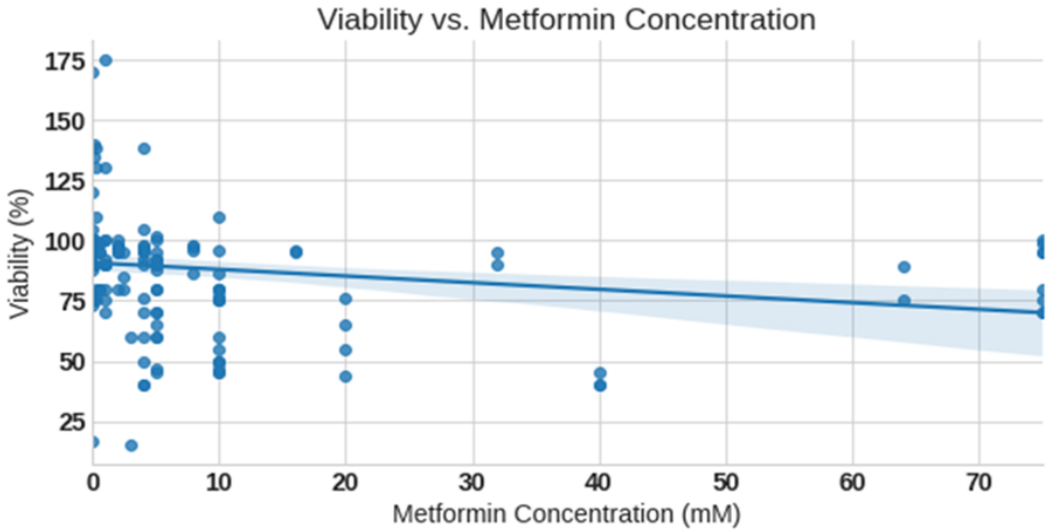

Figure 7: Simple linear regression model between viability (%) and metformin concentration (mM) collected from various published articles regarding metformin treatment of cancer cells.

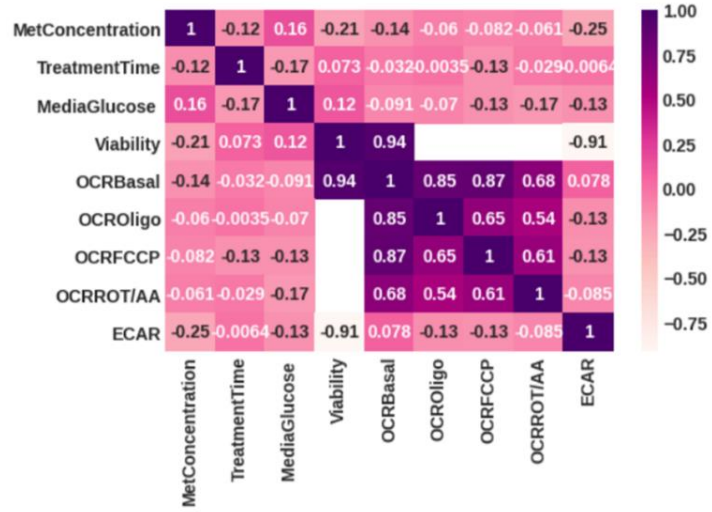

Figure 8: A correlation matrix developed using Python which displays the correlation coefficients between each variable in the database.

Support vector machine algorithms were applied after the data was divided into training and testing sets. As presented in table 3, the application of different SVM algorithms exhibited different accuracies depending on the kernel function. The application of the default linear kernel led to the highest accuracy of 86.9 %. Meanwhile, the usage of a polynomial kernel resulted in the lowest accurate model with 56.6 %. Algorithms with a default hyperparameter and RBF kernel displayed similar accuracies of ≈ 70%.

Table 3: The accuracy of support vector machine algorithms when applied to the database.

Support vector machine algorithms	**Accuracy of model**
Default hyperparameter	70 %
Default linear kernel	86.9 %
RBF kernel	69.7 %
Polynomial kernel	56.6 %

4 Discussion

4.1 Metformin Treatment of HCT116

After determining 5000 cells/well to be a suitable seeding density for the HCT116 cells, AlamarBlue and CCK-8 assays were performed on cells treated with MQ water and varying concentrations of metformin. For both these assays, two trials were performed and the cell viability was examined after 24 and 48 hours after treatment. As mentioned in the methods section, alamarBlue assay and CCK-8 measure viability in two different forms. While alamarBlue assay quantitatively measures metabolic activity, CCK-8 directly determines the number of viable cells. By taking this into consideration, the results of these assays can be better understood. In the first trial, a clear decreasing trend was observed in the 24-hour and 48-hour treatments of both assays (Fig.5). There appeared to be a correlation between decreasing metabolic activity/viability in cells and increasing metformin concentration. It was also apparent that the decrease in viability was at a greater extent in cells treated for a longer period. Unlike the gradual decrease after 24-hour treatments, the cells treated for 48 hours displayed a more pronounced decrease in viability. While viability was only reduced to ≤ 80% after 24 hours of 10 mM metformin treatment, 48-hour treatments reduced the viability to ≤ 60 %. Also, the viability

measured with CCK-8 assay after 48-hours was especially lower than that of alamarBlue assay. Based on this, it can be presumed that cancer cells treated with metformin may still exhibit somewhat high metabolic activity even though less viable cells are present.

Despite the decreasing trend in the first trial, there were some instances where cells treated with 5 mM and 10 mM displayed a slight increase in viability. However, these rises were marginal and presumably due to technical errors. There were possibly some variations in the number of cells seeded into the wells. This may have occurred when aspirating the medium before adding the treatments, as some cells may also have been removed from the wells, which could have resulted in more cells being present in some of the wells.

The second trial displayed somewhat varying trends, although a decreasing trend was observed in certain areas (Fig.6). There was a gradual decrease in viability in those treated with 1 mM and 5 mM for 24 hours. However, in both charts (Fig.6A and Fig.6B), the cells treated with 10 mM metformin exhibited higher viability than those treated with 5 mM. As this was primarily observed in the AlamarBlue assay, this increase does not necessarily mean that there were more viable cells in the 10 mM, but rather that these cells were exhibiting higher levels of metabolic activity.

The correlation between declining viability and increasing metformin concentration was more evident in the 48-hour treatments of the second trial. Yet, it was also observed that the cells treated with 1 mM metformin for 48 hours hardly seemed to be affected. In one instance, the cells treated with 1 mM metformin even exhibited higher viability than the untreated control cells (Fig.6A). This could imply that HCT116 cells are not very sensitive to metformin at lower concentrations. In the viability measured by performing a CCK-8 assay (Fig.6B), there appeared to be a more accelerated decrease in viability at higher concentrations of metformin. Here, a 10 mM treatment for 48 hours resulted in a viability of 35 %, which is considerably lower than the 60 % viability that was observed in the first trial.

4.2 Machine Learning

The negative linear relationship between metformin concentration and cellular viability was found using linear regression (Fig.7). As the correlation coefficient was found to be -0.21, it can

be established that this was a weak negative correlation. The graph seemed to be somewhat cluttered with numerous inconsistencies. These inconsistencies can be explained by the fact that the database included data from various cell lines of different organisms. From this, it can be presumed that cells of different origins display varying tolerance to metformin concentration. In the dataset, the highest metformin concentration was 75 mM, which is an unachievable dosage in humans (Table 2). If the data were organized according to the organism, perhaps a more representative model could be developed. This model could then be used to find possible treatments for humans. Furthermore, such models could also be used to find suitable drug combinations that would result in the best inhibition of cancer cell proliferation.

To find suitable SVM models that could be used to make accurate predictions, SVM algorithms with different kernels functions were applied. The polynomial kernel function led to the model with the poorest performance and lowest accuracy of 56.6 %. Also, while RBF kernel functions are often preferred, it exhibited slightly poor performance in this case with an accuracy of 69.7 %. These reduced performances may be due to the underfitting of the data. The database contained multiple gaps, which could have prevented these algorithms from perceiving a proper trend. It was the SVM algorithm with the default linear kernel that had the highest accuracy of 86.9 %. While this algorithm displays high performance, the model could be further improved by collecting larger samples of data. The variations in the performance of different SVM algorithms display how essential it is to pick the right kernel function to produce a model that can make predictions of high accuracy. From the high performance of the SVM algorithm with default linear kernel, it can be understood that machine learning has great potential to predict accurate drug responses of cancer cells.

5 Conclusion

5.1 Conclusion

By applying different machine learning algorithms to the database it was evident that predictions could be made regarding the response of cancer cells to metformin. From the weak negative correlation observed between cell viability and metformin concentration, it can be predicted how the cells would respond at higher concentrations of metformin. Support vector machine models also showed that they can be trained to make accurate predictions provided that the best-suited

kernel function is used. As the highest accuracy was 86.9 %, it is apparent that there is scope for improvement. It could especially be improved by adding larger amounts of data to the database, as this would enhance the clarity of the trend.

The effect of metformin on the viability of cancer cells observed in the dataset was also confirmed by exposing the colorectal cell line HCT116 to metformin. HCT116 established sensitivity to metformin treatment by displaying a decline in cell viability. It was notably evident that exposure to metformin for a longer period had a greater effect, as there were fewer viable cells present after 48-hours-treatments. Furthermore, there was predominantly a decreasing trend in viability as metformin concentrations were increased.

In conclusion, while the machine learning models displayed great potential to make predictions regarding cellular response, it was implied that various factors must be considered to ensure the development of an accurate model.

5.2 Future Perspectives

The implementation of machine learning models to predict the drug response of cancer cells can lead to formulations of new potential cancer treatments. It can be predicted how cell lines respond to individual drugs in comparison to a combination of drugs. The drug responses can also expand our knowledge regarding various mechanisms in cancer cell metabolism. As machine learning enables the usage of multiple variables, prediction models of greater accuracy can be developed. This can further be enhanced by combining multiple collections of databases. Going forward, more data can be collected as increasing numbers of studies are conducted. Subsequently, machine learning algorithms can be trained to create highly accurate models.

6 Publication bibliography

Adekola, Kehinde; Rosen, Steven T.; Shanmugam, Mala (2012): Glucose transporters in cancer metabolism. In *Current opinion in oncology* 24 (6), pp. 650–654. DOI: 10.1097/CCO.0b013e328356da72.

Agilent Technologies, Inc.: Seahorse XF Cell Mito Stress Test Kit User Guide. Available online at https://www.agilent.com/cs/library/usermanuals/public/XF_Cell_Mito_Stress_Test_Kit_User_Guide.pdf, checked on 11-Apr-21.

Alam, Md Maksudul; Lal, Sneha; FitzGerald, Keely E.; Zhang, Li (2016): A holistic view of cancer bioenergetics: mitochondrial function and respiration play fundamental roles in the development and progression of diverse tumors. In *Clinical and Translational Medicine* 5 (1), p. 3. DOI: 10.1186/s40169-016-0082-9.

Anderson, Nicole M.; Mucka, Patrick; Kern, Joseph G.; Feng, Hui (2018a): The emerging role and targetability of the TCA cycle in cancer metabolism. In *Protein & cell* 9 (2), pp. 216–237. DOI: 10.1007/s13238-017-0451-1.

Anderson, Rebecca G.; Ghiraldeli, Lais P.; Pardee, Timothy S. (2018b): Mitochondria in cancer metabolism, an organelle whose time has come? In *Biochimica et biophysica acta. Reviews on cancer* 1870 (1), pp. 96–102. DOI: 10.1016/j.bbcan.2018.05.005.

Binefa, Gemma; Rodríguez-Moranta, Francisco; Teule, Alex; Medina-Hayas, Manuel (2014): Colorectal cancer: from prevention to personalized medicine. In *World journal of gastroenterology* 20 (22), pp. 6786–6808. DOI: 10.3748/wjg.v20.i22.6786.

CloudFactory: The Essential Guide to Quality Training Data for Machine Learning. Available online at https://www.cloudfactory.com/training-data-guide, checked on 08-Apr-21.

DeBerardinis, Ralph J.; Chandel, Navdeep S. (2016): Fundamentals of cancer metabolism. In *Science Advances* 2 (5), e1600200. DOI: 10.1126/sciadv.1600200.

Dojindo, Cell Counting Kit-8 - Technical Manual.

Dong, Lanfeng; Neuzil, Jiri (2019): Targeting mitochondria as an anticancer strategy. In *Cancer Commun* 39 (1), p. 63. DOI: 10.1186/s40880-019-0412-6.

ECACC General Cell Collection: 91091005 HCT 116.

Erickson, Bradley J.; Korfiatis, Panagiotis; Akkus, Zeynettin; Kline, Timothy L. (2017): Machine Learning for Medical Imaging. In *Radiographics* 37 (2), pp. 505–515. DOI: 10.1148/rg.2017160130.

Ferreira. D; Adega. F; Chaves. R (2013): The Importance of Cancer Cell Lines as in vitro Models in Cancer Methylome Analysis and Anticancer Drugs Testing | IntechOpen. Available online at https://www.intechopen.com/books/oncogenomics-and-cancer-proteomics-novel-approaches-in-biomarkers-discovery-and-therapeutic-targets-in-cancer/the-importance-of-

cancer-cell-lines-as-in-vitro-models-in-cancer-methylome-analysis-and-anticancer-d, updated on 2013, checked on 13-Apr-21.

Ganapathy-Kanniappan, Shanmugasundaram; Geschwind, Jean-Francois H. (2013): Tumor glycolysis as a target for cancer therapy: progress and prospects. In *Molecular cancer* 12, p. 152. DOI: 10.1186/1476-4598-12-152.

Goel, Vishabh (2018): Building a Simple Machine Learning Model on Breast Cancer Data. In *Towards Data Science*, 2018. Available online at https://towardsdatascience.com/building-a-simple-machine-learning-model-on-breast-cancer-data-eca4b3b99fa3, checked on 01-Apr-21.

Huang, Cai; Mezencev, Roman; McDonald, John F.; Vannberg, Fredrik (2017): Open source machine-learning algorithms for the prediction of optimal cancer drug therapies. In *PloS one* 12 (10), e0186906. DOI: 10.1371/journal.pone.0186906.

Kalyanaraman, Balaraman (2017): Teaching the basics of cancer metabolism: Developing antitumor strategies by exploiting the differences between normal and cancer cell metabolism. In *Redox biology* 12, pp. 833–842. DOI: 10.1016/j.redox.2017.04.018.

Kalyanaraman, Balaraman; Cheng, Gang; Hardy, Micael; Ouari, Olivier; Lopez, Marcos; Joseph, Joy et al. (2018): A review of the basics of mitochondrial bioenergetics, metabolism, and related signaling pathways in cancer cells: Therapeutic targeting of tumor mitochondria with lipophilic cationic compounds. In *Redox biology* 14, pp. 316–327. DOI: 10.1016/j.redox.2017.09.020.

Lao, Victoria Valinluck; Grady, William M. (2011): Epigenetics and colorectal cancer. In *Nature reviews. Gastroenterology & hepatology* 8 (12), pp. 686–700. DOI: 10.1038/nrgastro.2011.173.

Liu, Chao; Liu, Qianqian; Yan, Aiwen; Chang, Hui; Ding, Yuyin; Tao, Junye; Qiao, Chen (2020): Metformin revert insulin-induced oxaliplatin resistance by activating mitochondrial apoptosis pathway in human colon cancer HCT116 cells. In *Cancer Medicine* 9 (11), pp. 3875–3884. DOI: 10.1002/cam4.3029.

Liu, Yu'e; Shi, Yufeng (2020): Mitochondria as a target in cancer treatment. In *MedComm* 1 (2), pp. 129–139. DOI: 10.1002/mco2.16.

Mirabelli, Peppino; Coppola, Luigi; Salvatore, Marco (2019): Cancer Cell Lines Are Useful Model Systems for Medical Research. Available online at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6721418/, updated on 2019, checked on 13-Apr-21.

Muse® Count & Viability Kit (200X) User's Guide (2020).

Nitzsche, Richard; Nishith Gupta (2017): Genetic dissection of the central carbon metabolism in the intracellular parasite Toxoplasma gondii.

Olawale, Ibrahim (2020): Linear Models in Practice - DSN AI+ FUTA - Medium. In *DSN AI+ FUTA*, 2020. Available online at https://medium.com/dsn-ai-futa/linear-models-in-practice-d47710854024, checked on 11-Apr-21.

Owen, M. R.; Doran, E.; Halestrap, A. P. (2000): Evidence that metformin exerts its anti-diabetic effects through inhibition of complex 1 of the mitochondrial respiratory chain. In *Biochemical Journal* 348 Pt 3 (Pt 3), pp. 607–614.

pandas.DataFrame. pandas 1.2.3 documentation (2021). Available online at https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.html, updated on 2021, checked on 11-Apr-21.

Porporato, Paolo Ettore; Filigheddu, Nicoletta; Pedro, José Manuel Bravo-San; Kroemer, Guido; Galluzzi, Lorenzo (2018): Mitochondrial metabolism and cancer. In *Cell research* 28 (3), pp. 265–280. DOI: 10.1038/cr.2017.155.

Santidrian, Antonio F.; Matsuno-Yagi, Akemi; Ritland, Melissa; Seo, Byoung B.; LeBoeuf, Sarah E.; Gay, Laurie J. et al. (2013): Mitochondrial complex I activity and NAD+/NADH balance regulate breast cancer progression. In *The Journal of Clinical Investigation* 123 (3), pp. 1068–1081. DOI: 10.1172/JCI64264.

Savas, Caner; Dovis, Fabio (2019): The Impact of Different Kernel Functions on the Performance of Scintillation Detection Based on Support Vector Machines. In *Sensors (Basel, Switzerland)* 19 (23). DOI: 10.3390/s19235219.

Sidey-Gibbons, Jenni A. M.; Sidey-Gibbons, Chris J. (2019): Machine learning in medicine: a practical introduction. In *BMC medical research methodology* 19 (1), p. 64. DOI: 10.1186/s12874-019-0681-4.

Song Yan-yan; LU, Ying (2015): Decision tree methods: applications for classification and prediction. In *Shanghai Archives of Psychiatry* 27 (2), pp. 130–135. DOI: 10.11919/j.issn.1002-0829.215044.

Sun, Xingming (Ed.) (2019): Artificial intelligence and security. 5th international conference, ICAIS 2019, New York, NY, USA, July 26-28, 2019 : proceedings. Cham: Springer (Lecture notes in computer science, 11634). pp. 395

Tan, Bie; Xiao, Hao; Li, Fengna; Zeng, Liming; Yin, Yulong (2015): The profiles of mitochondrial respiration and glycolysis using extracellular flux analysis in porcine enterocyte IPEC-J2. In *Animal Nutrition* 1 (3), pp. 239–243. DOI: 10.1016/j.aninu.2015.08.004.

ThermoFisher, alamarBlue® Assay.

Uddin, Shahadat; Khan, Arif; Hossain, Md Ekramul; Moni, Mohammad Ali (2019): Comparing different supervised machine learning algorithms for disease prediction. In *BMC medical informatics and decision making* 19 (1), p. 281. DOI: 10.1186/s12911-019-1004-8.

Urra, Félix A.; Muñoz, Felipe; Lovy, Alenka; Cárdenas, César (2017): The Mitochondrial Complex(I)ty of Cancer. In *Frontiers in Oncology* 7, p. 118. DOI: 10.3389/fonc.2017.00118.

Vancura, Ales; Bu, Pengli; Bhagwat, Madhura; Zeng, Joey; Vancurova, Ivana (2018): Metformin as an Anticancer Agent. In *Trends in pharmacological sciences* 39 (10), pp. 867–878. DOI: 10.1016/j.tips.2018.07.006.

Vial, Guillaume; Detaille, Dominique; Guigas, Bruno (2019): Role of Mitochondria in the Mechanism(s) of Action of Metformin. In *Front. Endocrinol.* 10, p. 294. DOI: 10.3389/fendo.2019.00294.

Weinberg, Samuel E.; Chandel, Navdeep S. (2015): Targeting mitochondria metabolism for cancer therapy. In *Nature chemical biology* 11 (1), pp. 9–15. DOI: 10.1038/nchembio.1712.

What Is Colorectal Cancer? | CDC (2021). With assistance of Division of Cancer Prevention and Control, Centers for Disease Control and Prevention. Available online at https://www.cdc.gov/cancer/colorectal/basic_info/what-is-colorectal-cancer.htm, updated on Feb 2021, checked on 27-Mar-21.

Wheaton, William W.; Weinberg, Samuel E.; Hamanaka, Robert B.; Soberanes, Saul; Sullivan, Lucas B.; Anso, Elena et al. (2014): Metformin inhibits mitochondrial complex I of cancer cells to reduce tumorigenesis. In *eLife* 3, e02242. DOI: 10.7554/eLife.02242.

Yang, Yuhui; Karakhanova, Svetlana; Hartwig, Werner (2016): Mitochondria and Mitochondrial ROS in Cancer: Novel Targets for Anticancer Therapy. Available online at https://onlinelibrary.wiley.com/doi/full/10.1002/jcp.25349, updated on 2016, checked on 19-Apr-21.

Yu, Li; Chen, Xun; Sun, Xueqi; Wang, Liantang; Chen, Shangwu (2017): The Glycolytic Switch in Tumors: How Many Players Are Involved? In *Journal of Cancer* 8 (17), pp. 3430–3440. DOI: 10.7150/jca.21125.

"HCT 116", *ECACC General Cell Collection: 91091005 HCT 116*, Culture collections, https://www.phe-culturecollections.org.uk/products/celllines/generalcell/detail.jsp?refId=91091005&collection=ecacc_gc, [Date of access 24.03.21]

7 Appendices

7.1 Coding – Support Vector Machines

```
[18] # Default hyperparameter
     from sklearn.svm import SVC
     from sklearn import metrics
     svc=SVC() #Default hyperparameters
     svc.fit(train_X,train_y)
     pred_y=svc.predict(test_X)
     print('Accuracy Score:')
     print(metrics.accuracy_score(test_y,pred_y))
```

```
Accuracy Score:
0.680327868852459
```

```
[21] # performing SVM by taking hyperparameter gamma=0.01 and kernel as rbf
     from sklearn.svm import SVC
     svc= SVC(kernel='rbf',gamma=0.01)
     svc.fit(train_X,train_y)
     predict_y=svc.predict(test_X)
     metrics.accuracy_score(test_y,predict_y)
```

```
0.7295081967213115
```

```
[22] # performing SVM by taking hyperparameter degree=3 and kernel as poly
     from sklearn.svm import SVC
     svc= SVC(kernel='poly',degree=3)
     svc.fit(train_X,train_y)
     predict_y=svc.predict(test_X)
     accuracy_score= metrics.accuracy_score(test_y,predict_y)
     print(accuracy_score)
```

```
0.6065573770491803
```

```
[23] # Default Linear kernel
     svc=SVC(kernel='linear')
     svc.fit(train_X,train_y)
     pred_y=svc.predict(test_X)
     print('Accuracy Score:')
     print(metrics.accuracy_score(test_y,pred_y))
```

```
Accuracy Score:
0.860655737704918
```

```
[24] # Default RBF kernel
     svc=SVC(kernel='rbf')
     svc.fit(train_X,train_y)
     pred_y=svc.predict(test_X)
     print('Accuracy Score:')
     print(metrics.accuracy_score(test_y,pred_y))
```

```
Accuracy Score:
0.680327868852459
```

Figure 7.1.1: The coding used to find the accuracy of various SVM models when applied to the database.

```
[8]  print(train.shape)
     print(test.shape)

     (282, 9)
     (122, 9)
```

```
[9]  #the variables which will be used for prediction
     prediction_var = ['MetConcentration', 'TreatmentTime', 'MediaGlucose ', 'Viability',
            'OCRBasal', 'OCROligo', 'OCRFCCP', 'OCRROT/AA', 'ECAR']
```

```
[10] train_X = train[prediction_var] #training data input
     train_y = train.Viability #ouput of training data

     test_X = test[prediction_var] #testing data input
     test_y = test.Viability #output of testing data
```

```
[17] model.fit(train_X,train_y) #fitting the model for training data

     SVC(C=1.0, break_ties=False, cache_size=200, class_weight=None, coef0=0.0,
         decision_function_shape='ovr', degree=3, gamma='scale', kernel='rbf',
         max_iter=-1, probability=False, random_state=None, shrinking=True,
         tol=0.001, verbose=False)
```

```
[16] test_y = np.nan_to_num(test_y)
     test_X = np.nan_to_num(test_X)

     train_X = np.nan_to_num(train_X)
     train_y = np.nan_to_num(train_y)
```

Figure 7.1.2: The coding used to divide the data into training and testing sets, before training the data.

7.2 Coding – Linear Regression and Correlation matrix

```
[28]
     sns.lmplot(x='MetConcentration', y='Viability',data=df,aspect=2,height=6)
     plt.xlabel('Metformin Concentration (mM)')
     plt.ylabel('Viability (%)')
     plt.title('Viability vs. Metformin Concentration')

     Text(0.5, 1.0, 'Viability vs. Metformin Concentration')
```

```
[30] # Correlation matrix showing correlation coefficients between each variable
     corr = df.corr()
     sns.heatmap(corr, cmap = 'RdPu', annot= True);
```

Figure 7.2.1: The coding used to develop a linear regression model and a correlation map.

Link to google colaboratory:
https://colab.research.google.com/drive/138p3cPiKBvRwsvcGGZZja_BwDXTPXGVK?usp=sharing