**Title (German)**

Subtitle (German)

**Title (English)**

Subtitle (English)

Bachelor Thesis

Submitted in partial fulfilment of the requirements for the degree of

Bachelor of

the University of Applied Sciences FH Campus Wien

Bachelor’s degree Program for Advanced Professional Training: Bioengineering

Author:

first name surname

Student identification number:

Number

Supervisor:

FH-Prof.in Dr.in Alexandra Graf

Norbert Auer, MSc

Date:

dd. mm. yyyy

Acknowledgements

(optional)

Abstract

Because on-line search databases typically contain only abstracts, it is vital to write a complete but concise description of your work to entice potential readers into obtaining a copy of the full paper. Writers should follow a checklist consisting of: motivation, problem statement, approach, results, and conclusions. Following this checklist should increase the chance of people taking the time to obtain and read your complete paper.

Kurzfassung

Dieser Teil ist gleich dem Abstract von oben nur in deutscher Sprache.

List of Abbreviations

DNA jfjfj

CSC jjjj

WHO ggg

Table of Contents

[Acknowledgements 2](#_Toc198260770)

[Abstract 3](#_Toc198260771)

[List of Abbreviations 4](#_Toc198260772)

[1. Introduction 7](#_Toc198260773)

[1.1. Aim 7](#_Toc198260774)

[1.2. Brain tumour – General 7](#_Toc198260775)

[1.2.1. Cancer stem cells 7](#_Toc198260776)

[1.2.2. Medulloblastomas 8](#_Toc198260777)

[1.2.3. Gliomas 8](#_Toc198260778)

[1.2.4. Therapeutic approaches 9](#_Toc198260779)

[1.2.5. Sub-subchapter 9](#_Toc198260780)

[1.3. Brain tumour – General 9](#_Toc198260781)

[1.3.1. Sub-subchapter 9](#_Toc198260782)

[1.3.2. Sub-subchapter 9](#_Toc198260783)

[2. Materials and Methods 10](#_Toc198260784)

[2.1. Hardware and System 10](#_Toc198260785)

[2.2. Dataset 10](#_Toc198260786)

[2.3. Python 10](#_Toc198260787)

[2.3.1. Library TensorFlow 10](#_Toc198260788)

[2.3.2. Library Keras 10](#_Toc198260789)

[2.3.3. Library Matplotlib 11](#_Toc198260790)

[2.3.4. Function Confusion matrix 11](#_Toc198260791)

[3. Results 12](#_Toc198260792)

[3.1. Visualisation of the MRI images 12](#_Toc198260793)

[3.2. Image settings and creation of CNN model 12](#_Toc198260794)

[3.2.1. Model with low accuracy: 13](#_Toc198260795)

[3.2.2. Model with very high accuracy 13](#_Toc198260796)

[3.3. Visual representation of accuracy and loss 13](#_Toc198260797)

[3.4. Confusion matrix 15](#_Toc198260798)

[4. Discussion 17](#_Toc198260799)

[5. Conclusion 17](#_Toc198260800)

[6. Bibliography 18](#_Toc198260801)

[6.1. Links 18](#_Toc198260802)

[7. List of Figures 19](#_Toc198260803)

[8. List of Tables 19](#_Toc198260804)

[9. Appendix 19](#_Toc198260805)

1. Introduction
   1. Aim

Blablabla

* 1. Brain tumour – General

There are around 150 different types of brain tumour, which either develop primarily directly in the brain or migrate into the brain through metastasising tumours from other tissues. A distinction is made between tumours in children and adults, which are very different in their development process and therefore the therapy must be specifically adapted. [1]

Paediatric brain tumours account for a quarter of childhood cancers and are primarily caused by epigenetic changes, such as DNA methylation and histone modifications, which regulate the expression of oncogenes and tumour suppressor genes. These changes also influence the formation of cancer stem cells and promote tumour growth. In contrast to adults, mutations occur rather rarely, but cell growth is faster. [1] [2]

Tumours in adults are mainly caused by genetic mutations and are influenced by environmental factors and the microenvironment. [1] [3]

Each type of brain tumour develops from several specific cells of origin, such as stem cells, progenitor cells or mature cells, so that different subtypes exist within a tumour group. Based on the different degree of differentiation, the location, specific markers and the source of the original cell in the cerebellum, tumours can be classified into subgroups. The cells of origin influence the subsequent behaviour of the tumour and react with varying degrees of sensitivity to genetic mutations, which is crucial for tumour growth, prognosis, treatment methods and risk of relapse. Not every mutated stem cell transforms directly, but first differentiates into limited precursor cells. Also, more mature cells are more resistant to transformation than stem or progenitor cells. [1] [4] [5]

In addition to the cell of origin, there is also the so-called “cell of mutation”, which is not immediately transformed, but is the first cell to contain a relevant genetic change that activates certain tumour-forming signalling pathways. [5]

Tumour grading from grade 1 to 4 of brain tumours is based on histopathological, clinical and molecular findings according to the World Health Organization (WHO) “Blue Book”, fifth edition of 2021. Grade 1 are slow-growing, benign tumours, while grade 4 includes highly heterogeneous, fast-growing, usually fatal tumours. [6] [7]

* + 1. Cancer stem cells

The development of the nervous system is based on a balance between the self-renewal and differentiation of neural stem cells, which is highly precisely regulated in the brain. However, if this process is disrupted, for example by mutations, this can lead to the formation of subpopulations of tumour cells that have stem cell-like properties and contribute to the development of brain tumours. These cells are called cancer stem cells (CSC) and block the normal differentiation of progenitor cells and cause uncontrolled proliferation, leading to tumour growth, resistance to therapy and relapse. Stem cells are the most frequent tumour initiators, but differentiated progenitor cells can also acquire stem cell-like properties through genetic or epigenetic mutations. If this leads to dedifferentiation or to an active blockade of dedifferentiation, the cells enter a proliferative state and form the basis for tumour development.

A special characteristic of CSCs is their resistance to aggressive treatments such as radiotherapy or chemotherapy. The reason for this is their ability to survive through improved DNA repair processes, activation of checkpoints and support of the tumour microenvironment. In addition, CSCs are able to regenerate cell types from the heterogeneous tumour mass after therapy by reactivating tumour-relevant signalling pathways and activating genes that promote cell self-renewal, which often leads to relapse. [4] [3]

The tumour microenvironment also plays an important role, where the cancer stem cells are located in specific hypoxic niches and are not randomly distributed in the tumour. This is where neuronal activities and interactions with immune cells take place, which contribute to the development and maintenance of the tumour. [3] [2]

* + 1. Medulloblastomas

Medulloblastomas are among the most common and most dangerous paediatric brain tumours, which can usually only be treated with a combination of radiotherapy and chemotherapy and therefore have severe side effects. They are classified as WHO grade 4 and can be categorised into four molecular groups based on their site of origin. [4] [3]

Group 1 is a desmoplastic subtype that arises from granulocyte precursor cells in the cerebellum and is controlled by the Sonic Hedgehog signalling pathway. In contrast, group 2 is a classic subtype originating from progenitor cells in the dorsal brainstem and is regulated by the WNT signalling pathway. Group 3 is characterised by Myc overexpression in various progenitor cells and group 4 has cells of origin from the superior rhomboid lip. These four main groups can be subdivided into a further twelve subgroups that require individual therapy procedures in order to treat the tumour in the best possible way. [4] [5]

* + 1. Gliomas

Gliomas are among the most common types of brain tumours and resemble normal glial cells. They are divided into neural, proneural, classic and mesenchymal subtypes, which can have a WHO tumour grade of 1 to 4. There are two different approaches to their development. On the one hand, gliomas can develop from neuronal stem cells, for example if mutations occur, and on the other hand, oligodendrocyte precursor cells are partly responsible for tumour growth as they are susceptible to genetic changes. [5]

The most dangerous of these subtypes is glioblastoma, which has a WHO grade 4 and is often lethal even when treated. [4] [3]

* + 1. Therapeutic approaches

The treatment of brain tumours is very individual and complex. It is therefore usually difficult to find the correct therapy. Children often respond better to treatment than adults, but the side effects can be more intense and have long-term consequences. In adults, however, the risk of metastasis and recurrence is higher, leading to worse prognoses. [1] Due to the presence of therapy-resistant, reactive cancer stem cells, radiotherapy and chemotherapy only have a palliative effect, even when used repeatedly, and make treatment even more difficult. [3] In addition, the blood-brain barrier impedes access for chemotherapeutic agents and other anti-tumour drugs. To overcome this, the invasive variant directly attacks the blood-brain barrier and thus opens it. There is also a non-invasive method in which nanoparticles containing the active substance are transported through receptors. [1]

Knowledge of the genetic and epigenetic properties, as well as the cell of origin and molecular mechanisms of tumour development, forms the basis for effective therapies. The long-term goal is a personalised treatment that enables a direct and effective attack on the tumour cells in order to spare the surrounding healthy tissue. [3] [4] This can be achieved by combining standard therapies with immunotherapies that block various signalling pathways, destroy cancer stem cells or increase their sensitivity to chemotherapy. The correct dosage and sequences play a decisive role here. [1] [3]

Furthermore, information about the tumour microenvironment is relevant to enable targeted changes and improve the attack on hypoxic niches, which can contribute to treatment success. [3]

* + 1. Sub-subchapter

Text with additional image or table, if required.

* 1. Magnetic resonance imaging

Blablabla

* + 1. Sub-subchapter

Blablabla

* + 1. Sub-subchapter

Text with another equation. Commas have to be placed correctly.

1. Materials and Methods

This section contains details on the data used as well as Python and the additionally installed libraries

* 1. Hardware and System

The code was written on an HP notebook 15-ay589ng, which has an Intel(R) Core(TM) i3-6006U processor and runs on the Microsoft Windows 10 Home operating system.

* 1. Dataset

The dataset “Brian Tumor Dataset” was downloaded from the website kaggle on 8th March 2025 and includes 2513 MRI images of the Brain Tumour class (55%) and 2087 MRI images of the Healthy class (45%). They are divided into two separate folders according to their class. It also contains a CSV-file containing the metadata of the individual image data. [8]

The path to the folders was defined in Python to enable access to the image data.

* 1. Python

The Python programming language was first published in 1991 and is a freely available, open-source software for all platforms. By using indentation instead of curly brackets, it is very easy to read and well structured. Python also contains an extensive standard library and supports a large number of additional packages. It is well suited for machine learning to create, train and improve models. [9], [10]

For this work Python version 3.12.5 was used and the code was created in a Jupyter notebook. The Visual Studio Code software was used for this purpose, where the commands were written, edited and tested directly line by line.

The libraries os, TensorFlow and matplotlib.pyplot were also installed, so that the commands could be executed. In order to create the confusion matrix, the ConfusionMatrixDisplay function was imported from the sklearn.metrics library.

* + 1. Library TensorFlow

The TensorFlow library has been publicly available since November 2015 and was developed by Google to support deep learning applications. Although it was originally written in the C++ language, it is most commonly used in Python, among other things to classify images. TensorFlow is used to create models from neural networks. The image data (tensors) are processed at each network point (neuron) and flow to the next point via the network lines. This allows the generation of algorithms and the creation of a calculation graph. [11]

* + 1. Library Keras

Keras has been an extension library and part of Tensor Flow since 2017, which is why it does not need to be installed separately. This makes it easier to create calculation graphs and also means that fewer lines of code are required for the model. [11]

It also contains regularisation techniques such as data augmentation, where the training data set is artificially increased by rotating, reflecting and slightly shifting the images. [12]

For this work, Keras was used to develop the sequential CNN model of three consecutive stacks of convolutional layer and pooling layer followed by a fully connected layer. First, the images were normalised and certain parameters adjusted for data augmentation using the ImageDataGenerator function and split into 80% training data and 20% validation data. Then the model was created and the final settings for the model training were defined, such as the image size and the number of filters in each convolutional layer. Finally, the data were ready for the model and the number of epochs was defined for how often the entire training data set should be run through. To avoid overfitting, the Keras function EarlyStopping was used, where the training is stopped after a certain number of epochs as soon as the prediction accuracy of the validation data is no longer increased. [12]

Due to the high number of images, training the model is very time-consuming, but necessary to achieve a very high level of accuracy so that it can make the right predictions in the future.

* + 1. Library Matplotlib

The Matplotlib library is used to visualise data in a simple and quick way. It can be used to create a variety of diagrams including their axes, such as line charts, bar charts, matrices or box plots. It can also be used to visualise images.

Matplotlib.pyplot was used to visualise the first MRI images, to display the plots for the accuracy and loss function and to create the confusion matrix. [11], [13]

* + 1. Function Confusion matrix

In order to determine the quality of the predictions of this model, a confusion matrix was created that compares the predictions with the correct target values. This was used to determine the false positive and negative as well as the correct positive and negative predictions and to analyse whether this model is a good fit or not. [12]

1. Results  
   1. Visualisation of the MRI images

To check whether the path accesses the correct folders, the first image of each of the two subfolders “Brain tumour” and “Healthy” was displayed visually (Figure 1)

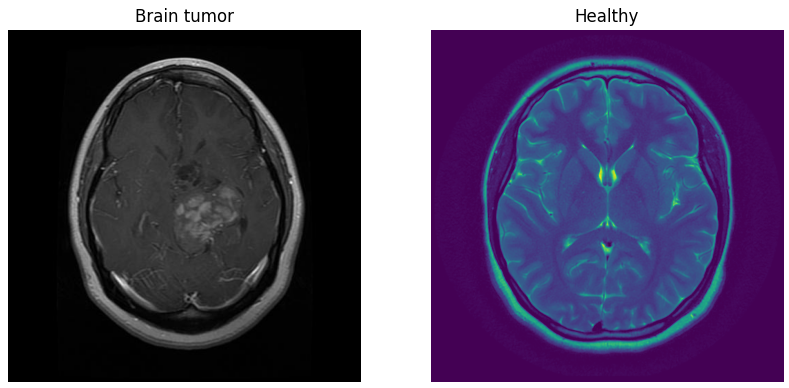


Figure 1: On the left is an MRI image of a patient with a brain tumour. In comparison, on the right is an image of a healthy patient

* 1. Image settings and creation of CNN model

For the CNN model, several runs of the code were necessary to find the appropriate parameters. Among other things, settings for data augmentation such as rotation, mirroring and shifting were first adjusted and then a different number of epochs were tried out.

The CNN model in general was created from three consecutive blocks of convolutional layer and pooling layer followed by a fully connected layer. The image size was 150x150 pixels with 32 channels at the start and 17x17 pixels with 128 channels at the output.

Figure 2 shows which layer is involved in the first column and the corresponding values in the second column. The width and height of the image are reduced by half in the pooling layers, while the number of channels is doubled. Flatten creates a 1D vector with 36992 entries, which results from the multiplication of height, width and channels of the last pooling layer. Dense\_2 shows the complete connected layer with 128 channels and the last layer gives the output for a binary model with probability 0 or 1.



Figure 2: Column 1 shows the individual layers and column 2 their corresponding parameters

To illustrate the importance of these factors, the model with very high accuracy was compared with a model with low accuracy.

* + 1. Model with low accuracy:

In order to artificially enlarge the image data set, the parameters for rotation, mirroring, zoom and horizontal / vertical shift were initially set to 10% each. For the training data, however, this only resulted in an accuracy of 86.85% and a loss of 30.23% for 10 epochs. For the validation data, the accuracy was only 79.54% and the loss 45.44%.

With 20 epochs, but constant parameters, an accuracy of 94.82% was achieved for the training data and a loss of 13.44%, for the validation data an accuracy of 83.57% and a loss of 39.97% was achieved.

* + 1. Model with very high accuracy

The images in this model were previously only normalised, but this time no data augmentation was performed. This resulted in a very high accuracy for the training data of 99.93% and a loss of 0.66% for 10 epochs. For the validation data, the accuracy was 98.80% and the loss 3.04%.

To check the stability of the model, the code was run four times with unchanged settings and an accuracy of >99% was achieved for each run.

* 1. Visual representation of accuracy and loss

In order to compare the accuracy and loss of the models, these were visualised graphically and compared.

The plots show that accuracy and loss behave in exactly the opposite way. The higher the accuracy, the lower the loss.

Figure 3 and Figure 5 show the results from the model with low accuracy and high loss, where the graph of the validation data set shows very strong fluctuations.

In comparison, the results from the very high accuracy and very low loss model are stable and there is only one peak at epoch 5, shown in Figure 4 and Figure 6.

|  |  |  |
| --- | --- | --- |
| Figure 3: Accuracy training and validation of model with low accuracy |  | Figure 4: Accuracy training and validation of model with very high accuracy |
|  |  |  |
| Figure 5: Loss training and validation of model with low accuracy |  | Figure 6: Loss training and validation of model with very high accuracy |

* 1. Confusion matrix

The confusion matrix of the worst model run was compared here with that of the best run. Figure 7 shows that 429 images of healthy subjects were identified as true negatives and 352 images with tumours were identified as true positives. However, 73 images were falsely identified as tumour positive, although they show healthy images, and 65 images were falsely identified as negative, although they show a tumour.

In the model with the highest accuracy, Figure 8 shows that 499 healthy images were classified as true negatives and 416 diseased images were classified as true positives. Only three images were incorrectly identified as tumours and a single tumour image was identified as falsely healthy.

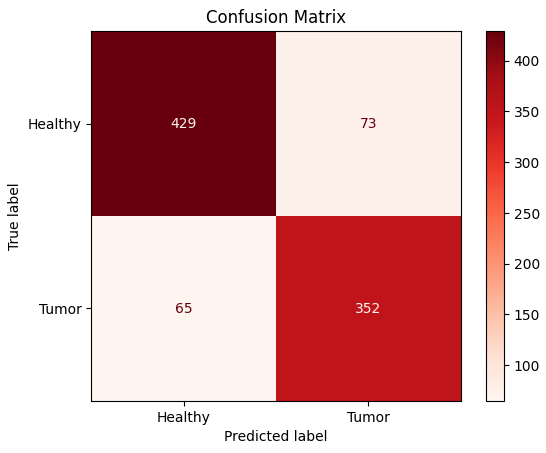


Figure 7: Confusion matrix of worst model

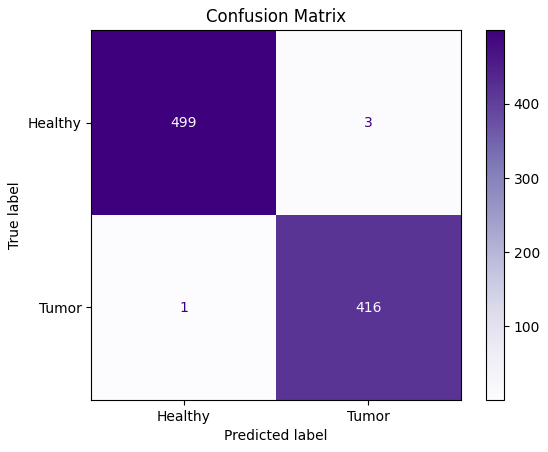


Figure 8: Confusion matrix of best model

Figure 9 shows the four matrices of the model runs without data augmentation, each of which has an accuracy of > 99% and thus confirms the robustness of the model.

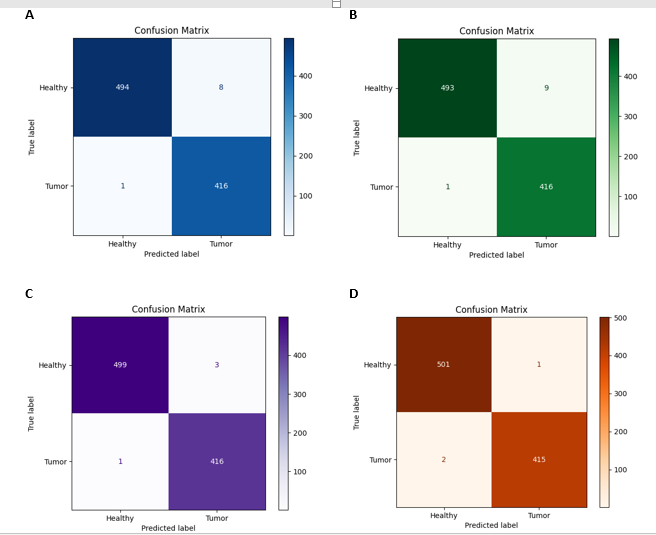


Figure 9: Confusion matrices of the best model for four runs to show the stability of the model

1. Discussion

Start with a few sentences that summarize the most important results. The discussion section should be a brief essay in itself, answering the following questions and caveats:

* Interpret results in terms of background laid out in the introduction - what is the relationship of the present results to the original question?
* How did you implement the question?
* What were the difficulties during the process?

1. Conclusion

What is the strongest and most important statement that you can make from your work?

* If you met the reader at a meeting six months from now, what do you want them to remember about your project?
* Refer back to problem posed and describe the conclusions that you reached from carrying out this work.
* Do not repeat word for word the abstract, introduction or discussion.

1. Bibliography

[1] S. Sarkar, T. Deyoung, H. Ressler, and W. Chandler, “Brain Tumors: Development, Drug Resistance, and Sensitization–An Epigenetic Approach,” 2023, *Taylor and Francis Ltd.* doi: 10.1080/15592294.2023.2237761.

[2] S. Marino and R. J. Gilbertson, “Harnessing brain development to understand brain tumours,” *Development (Cambridge)*, vol. 148, no. 2, Jan. 2021, doi: 10.1242/dev.193342.

[3] T. J. Abou-Antoun, J. S. Hale, J. D. Lathia, and S. M. Dombrowski, “Brain Cancer Stem Cells in Adults and Children: Cell Biology and Therapeutic Implications,” Apr. 01, 2017, *Springer New York LLC*. doi: 10.1007/s13311-017-0524-0.

[4] R. Azzarelli, B. D. Simons, and A. Philpott, “The developmental origin of brain tumours: A cellular and molecular framework,” May 01, 2018, *Company of Biologists Ltd*. doi: 10.1242/dev.162693.

[5] C. Liu and H. Zong, “Developmental origins of brain tumors,” Oct. 2012. doi: 10.1016/j.conb.2012.04.012.

[6] S. H. Torp, O. Solheim, and A. J. Skjulsvik, “The WHO 2021 Classification of Central Nervous System tumours: a practical update on what neurosurgeons need to know—a minireview,” *Acta Neurochir (Wien)*, vol. 164, no. 9, pp. 2453–2464, Sep. 2022, doi: 10.1007/s00701-022-05301-y.

[7] M. K. H. Khan *et al.*, “Machine learning and deep learning for brain tumor MRI image segmentation,” Nov. 01, 2023, *Frontiers Media SA*. doi: 10.1177/15353702231214259.

[8] “Brian Tumor Dataset.” Accessed: Mar. 08, 2025. [Online]. Available: https://www.kaggle.com/datasets/preetviradiya/brian-tumor-dataset

[9] “What is Python? Executive Summary | Python.org.” Accessed: Apr. 26, 2025. [Online]. Available: https://www.python.org/doc/essays/blurb/

[10] “Python: Definition und Verwendung der Programmiersprache.” Accessed: Apr. 26, 2025. [Online]. Available: https://www.bluebranch.de/lexikon/python

[11] T. Hope, Y. Resheff, and I. Lieder, *Einführung in TensorFlow : Deep-Learning-Systeme programmieren, trainieren, skalieren und deployen*. O’Reilly, 2018.

[12] Géron A., *Praxiseinstieg Machine Learning mit Scikit-Learn und TensorFlow : Konzepte, Tools und Techniken für intelligente Systeme*. o’Reilly, 2018.

[13] “matplotlib.pyplot — Matplotlib 3.5.3 documentation.” Accessed: Apr. 26, 2025. [Online]. Available: https://matplotlib.org/3.5.3/api/\_as\_gen/matplotlib.pyplot.html

* 1. Links

https://www.kaggle.com/datasets/preetviradiya/brian-tumor-dataset

<https://www.python.org/doc/essays/blurb/>

https://www.bluebranch.de/lexikon/python

https://matplotlib.org/stable/api/pyplot\_summary.html#module-matplotlib.pyplot

1. List of Figures

Figure 1: On the left is an MRI image of a patient with a brain tumour. In comparison, on the right is an image of a healthy patient 10

Figure 2: Column 1 shows the individual layers and column 2 their corresponding parameters 11

Figure 3: Accuracy training and validation of model with low accuracy 12

Figure 4: Accuracy training and validation of model with very high accuracy 12

Figure 5: Loss training and validation of model with low accuracy 12

Figure 6: Loss training and validation of model with very high accuracy 12

Figure 7: Confusion matrix of worst model 13

Figure 8: Confusion matrix of best model 13

Figure 9: Confusion matrices of the best model for four runs to show the stability of the model 14

1. List of Tables

Table 1: Title of table 1 3

Table 2: Title of table 2 3

1. Appendix

(Optional: Author’s appendix)

(Here diagrams, program details, calculation examples, statistics etc. can be inserted in separate sections.) NICHT DAS GESAMTE PROGRAMM! - Dieses Bitte auf die Moodle Plattform laden.