Klicka här för att ange ditt ämnesområde. T.ex. *Examensarbete inom teknik och lärande*)

Klicka här för att ange om exjobbet är på grundnivå eller avancerad nivå samt hur många högskolepoäng. T.ex. *Avancerad nivå, 15 hp*

Klicka här för att ange titel

Klicka här för att ange underrubrik

KLICKA HÄR FÖR ATT ANGE FÖRFATTARE (FÖR- OCH EFTERNAMN)

This is the English title

VILMER JONSSON
TOR STRIMBOLD

Abstract

English abstract goes here.

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Sammanfattning

abstract på svenska.

Contents

Chapter 1

Introduction

The frequency of malignant melanoma has in the last decade been on the rise with approximately 60 000 people being diagnosed with the disease yearly [1]. A similar trend has been observed in the United States where the frequency of people diagnosed with the disease has doubled yearly from 1982-2011 [2].

Skin cancer, although common, has a survival rate of up to 95% if discovered at an early stage. This fact makes the diagnosis of malignant melanoma an essential part of the treatment of the disease. In the light of diagnosis, machine learning has become a subject of interest in many medical fields, dermatology included. Multiple studies have been conducted to evaluate the potential of the algorithms' accuracy and applicability through so called computer aided diagnostics (CAD). Concerning malignant melanoma, some studies have reached an accuracy of up to 90% which competes with the experts in the field.

The accuracy of machine learning algorithms is however heavily reliant on the feature extraction methods used, where relevant features in an image are extracted and then used in the classification of the skin lesion. An effective feature extraction in turn relies on determining which features to use. Which is the subject of feature selection where different features are compared in terms of their effect on the diagnostic accuracy. It has been shown that effective feature selection make classifiers more effective, accurate and cost-effective. [3]

1.1   Research Question

In our thesis we will study the effect of two different feature selection methods upon 169 features and using 4 different machine learning models. We aim to answer:

1. Which features and combinations of features are the most important in terms of classification accuracy?

2. Is there a difference between different models concerning which features are most effective?

3. Are some feature selection methods more efficient in creating better accuracy than others? Which one uses the fewest number of features?

1.2   Approach

To study the proposed research questions both forward and backwards selection will be performed. To further widen the base for comparison four different machine learning models will be used: support vector machine, random forest, k-nearest neighbors and a one-layered neural network. These models are widely used in the research of automatic skin lesion diagnosis which makes them relevant to study. Furthermore, two selection methods will be used, namely sequential forwards selection and sequential backwards selection.

The features to be used in the feature selection are of two classes: hand-crafted and automatically generated features. The hand crafted features are based on the clinical practice through the ABCD method where a skin lesion is classified based on it's asymmetry, border irregularity, color and diameter. The ABCD method is widely used when diagnosing skin lesions using ML due to its effectiveness and simplicity to implement. [4] Thus making it relevant to the study. For the automatically generated features, SIFT was chosen as the feature.

Lastly the base of comparison will be widened by the use of two different datasets will be used in the study, namely the HAM10000 and ISIC dataset.

A comparison between the classifiers, where first no selection is done and then between the two selection methods, will be conducted. The result from this will then highlight whether or not different features have different impact on the efficiency and accuracy of machine learning algorithms. Furthermore, it will show if certain features are to be preferred in the context of different machine learning models. Lastly, it will show if there is any difference between the hand-crafted and automatically generated features.

1.3   Scope

The study is limited by the number of features, selection methods and models being used.

Firstly the models are limited by the fact that they consist of only machine learning. An extension of the study could include a deep learning model such as a convoluted neural network.

Secondly, the study is limited by the feature selection methods being used where only feature selection by wrapper methods are used through the sequential backward and forward selection.

Lastly, the study is limited by the features which the selection is based in three different ways. Firstly, not all combinations of the features will be examined since this would require too many computations. Secondly, the different aspects of the ABCD method can be extracted in a plethora of different ways, which leaves room for further studies. Thirdly, the features are only of two types of classes, other classes such as bag-of-features will not be used.

Chapter 2

Background

2.1   Skin Cancer

During the last decades the frequency of malignant melanoma in Sweden has been on the rise and the trend does not seem to be going down. Yearly, approximately 60 000 persons are diagnosed with skin cancer and around 500 die from the disease. [1]

Likewise in the United States, skin cancer is the most common type of cancer to the degree that every fifth american will develop the disease at some time during their lives. The rates of melanoma have on a general trend increased, nearly doubling between 1982-2011. On some levels the rates have decreased, as in for people that are under 30. [2]

Malignant melanoma is a type of skin cancer that is the result of melanocyte cells in the skin starting to grow uncontrollably. Although melanoma is a less common type of skin cancer, it is at the same time more prone to spreading to other parts of the body therefore making it more dangerous. [2]

Melanoma skin cancer has five different stages, stage 0-5, where in stage 0 the cancer is contained within the top layer of the skin while in the final stage the cancer has spread to other parts of the body [5]. According to UK researchers the survival rate of patients diagnosed with melanoma differs greatly depending on when during the four different stages the diagnosis is made [6]. Where if the cancer is detected in the first stage the survival rate is almost 100%, however if the diagnosis is done in a later stage the survival rate drops to 30%, although the statistic does not take into account the age of the patients. [6]

Recently artificial intelligence (AI) and machine learning (ML) has made great strides within dermatology in areas such as melanoma detection and clas-

sification. In one study, a machine learning algorithm was compared against experts, where the ML system outperformed the experts. [7]

2.2 Machine Learning

Machine learning (ML) is a field of computer science focusing on automated and intelligent data analysis. A ML algorithm is fed problems and answers whereupon the system tries to predict the answer. Based on the response the algorithm then makes adjustments to better give the correct answers. [8]

In the case of skin cancer detection, an ML algorithm would receive images with a ground truth for each image, meaning either the skin lesion is malignant or benign. The algorithm would then try to correctly decide if the lesion is cancerogenous thereafter the correct answer would be revealed and the algorithm makes corrections based on whether or not it was successful in the diagnosis.

An ML algorithm consists of four main steps: 1) pre-processing, 2) segmentation, 3) feature extraction and 4) classification. All of the named steps are described below.

2.2.1 Pre-processing

The main goal with pre-processing an image is to improve the quality of the picture. Improvement can be made in many ways but generally the methods can be sorted into either artifact removal or image enhancement [9]. Artifact removal entails removing misleading objects in the image such as hair pixels and air bubbles which can lead to incorrect segmentation and feature extraction. Image enhancement usually consists of contrast Pre-processing is an essential step in the process since the classification is vastly dependent on the image quality. [10]

2.2.2 Segmentation

In the segmentation phase the borders of the lesion are identified. This is done to be able to identify what in the image contains the lesion. Like the pre-processing, segmentation is vital for the classification and feature extraction since parameters such as the lesion's border cannot be correctly evaluated if the segmentation is poorly executed. At the same time, the segmentation of skin lesions is one of the most challenging steps due mainly to four reasons.

Firstly, most lesions do not have a stark contrast between the healthy and un-healthy skin. Secondly, the lesion's border may be irregular which impedes the segmentation. Thirdly, the color variegation within the lesion can be vast where multiple colors are present at the same time. Finally, there may be ar-tifacts such as hairs or lens flare present within the image which hinders the segmentation, this is however meant to be removed during the pre-processing. [10]

2.2.3   Feature Extraction

Clinical classification is done based on the characteristics of the lesion. Like-wise, the features of the lesion need to be quantified in a way that a computer can understand to make a classification. This is the goal of feature extraction, to quantify the relevant characteristics of the lesion. Which features to con-sider relevant depends on the base of classification, the same holds for the way to extract the features.

ABCD

The ABCD-rule is a mimic of the way clinicians determine whether or not a skin lesion is malignant or not based on four features: asymmetry, border, colour and dermoscopic structures or diameter of the lesion. The higher the asymmetry of the lesion, in terms of either shape or colour, the more likely it is to be malign. The border aspect refers to whether the border of the lesion is irregular. The lesion is divided into eight slices and for each slice, if the border is irregular a point is awarded. Regarding the colour, the lesion is scanned for the occurrence of six colours (white, red, light brown, dark brown, blue-grey and black). For each colour a point is awarded. The "D" in ABCD can either be evaluated using the dermoscopic structure or the diameter of the lesion. The dermoscopic structure is evaluated based on the presence of five structures. For each structure a point is awarded, so the values range from zero to five. If instead the diameter is evaluated, all lesions with a diameter greater than 6 mm are considered malignant and given a value of five. [11] [12]

Scale-invariant feature transform (SIFT)

Scale-invariant feature transform (SIFT) is a popular method to extract key points out of an image. These key points are detected by an algorithm and given descriptors based on why the points were considered important. These descriptors can then be used in a classifier.

SIFT features are, as previously mentioned, used widely in the broader image analysis field. However, the research where the features are used to classify melanoma is quite limited.

2.3   Classification

The classification of the lesions based on the features extracted can be done through several machine learning algorithms. The ones used in this study are listed below.

2.3.1   Support Vector Machine (SVM)

Support vector machines (SVM) are common and effective ML algorithms first introduced in the 1990s. SVM is an extension of the support vector classifier, which in turn is an extension of the maximal margin classifier.

The maximal margin classifier aims to separate data into two classes using a hyperplane. The support vector classifier builds on the same principle, but is more adjusted to the case when the data is not easy to separate. The SVM enlarges the feature space of the data in order to be able to separate data in a non-linear fashion. [13]

2.3.2   Random Forest (RF)

Random forest is a tree-based method. The model consists of several decision trees taking different factors into consideration. When classifying, all individual trees try to classify the data. Then, the class that most of the trees predicted is chosen as the final classification.

2.3.3   K-Nearest Neighbors (KNN)

K-nearest-neighbors is another method for classifying. The ground principle is to find the nearest neighboring data points from the training set for all the test data points. In a similar fashion to random forest, it performs a majority voting and then chooses the most occurring class from neighbors as the final prediction.

2.3.4   Neural Network (NN)

Neural networks are a type of machine learning algorithm that is inspired by the human brain. The network consists of several layers of neurons, where each neuron is connected to all neurons in the previous layer. The first layer is the input layer, where the data is fed into the network. The last layer is the output layer, where the final prediction is made. The layers in between are called hidden layers. The hidden layers are the ones that do the actual work of the network. The neurons in the hidden layers are connected to all neurons in the previous layer, and the output of each neuron is calculated by a function. The output of the neurons in the last hidden layer is then fed into the output layer, where the final prediction is made.

2.4   Feature selection

Feature selection is a process where the different extracted features are ranked in order of importance for the results of the classification. Some features may be irrelevant and thus confusing for the model. The goal for feature selection may vary, but it is commonly used to find the most important features, features that harm the results and synergies between features. In short, reducing the number of features utilized, improves performance and accuracy.

In this study the greedy methods Sequential Backward Elimination (SBE) and Sequential Forward Selection (SFS) will be used.

Sequential backward selection is a process that involves multiple iterations in which a feature is eliminated during each cycle. In the initial iteration, the feature that results in the worst performance, according to some metric, when removed is excluded from the set. The process is then repeated with the updated set of features. This process is continued until the desired number of features are left.

Forward selection shares the same underlying concept as backward selection, but instead begins with a smaller feature set and adds one feature during each iteration, resulting in an increased set of features rather than a reduced one.

2.5   Related Works

The automatic classification of skin lesions as being either malignant or benign has been a subject of great interest and there is a plethora of publications

investigating the manner. Hence, there are several studies focused around classifying lesions, using the feature set and ML models which this thesis aims to investigate. However, there is not such an abundant pool of studies which focus on feature selection methods and exploration of which types of features are the most important.

Första studied the minimal number of necessary features for effective and accurate skin lesion classification using a KNN-model. The features that were studies were based on the geometric (A), color (C) and boundary (B) aspects of the ABCD rule of melanoma.(Första) A feature selection method which implemented the Sequential Backward Selection (SBS) was used. When all features (15 in total) were used, the average accuracy was around 78%. The highest accuracy of 91% was obtained when there were only 6, 8 and 9 features. Hence, Första showed that an effective feature selection method both improves the accuracy of the classifier as well as the computation and storage cost. Though they did not highlight which of these features were abundant and which were important to include. Additionally they only limited the study to one type of ML model, namely the KNN.

Tre studied the effect of feature selection using the NCA method with an M-SVM on three different data sets. Like Första their method showed an increase in accuracy and runtime. However, as well as Första, they did not investigate which features were the ones leading to the increased accuracy. Furthermore, only one classifier was used with the M-SVM and therefore no results were generated using some other type of classifier.

Fyra used ECNA feature selection to gain succesful results. The highest accuracy reached was 94.73% using a KNN and second best was using an SVM with and accuracy of 93.83%. They however, used deep features instead of the hand crafted features which will be used in this study.

Chapter 3

Methods

3.1 Machine learning methods

3.1.1 Preprocessing

For removing artifacts such as hairs, the DullRazor® software was used. The software is created by Tim Lee, Vincent Ng, RIchard Gallagher, Andrew Coldman and David Mclean and is used frequently for preprocessing of skin lesion images. The method removes the hairs in two steps: 1) using generalized grayscale morphological closing operations, it identifies the dark hair locations, 2) verifies the shape of the hair pixels as long and thin, whereafter it replaces the verified pixels by bilinear interpolation and lastly 3) smooths the replaced hair pixels.

3.1.2 Segmentation

The aim of this study was not to develop a new segmentation method, but rather to compare the effect of different feature selection methods on the accuracy of the classification. Therefore, the segmentation was done using open-source software published by [name] availible on GitHub.

3.1.3 Feature extraction

As mentioned in the background, the two feature extraction methods used were ABCD and SIFT.

ABCD

The ABCD features were extracted using the open-source software published by [name] availible on GitHub. The software extracts the features based on the method proposed in [this] study.

To calculate the asymmetry and border features the measure library from sci-kit is used, with the `get_region_props` function. The color features are calculated using the color library from sci-kit.

SIFT

The SIFT features are extracted using the OpenCV class SIFT which implements the algorithm by David G. Lowe. After that...

3.1.4 Classification

All classifiers used in this study are from the Scikit-sklearn library which is a free-to-use library for machine learning in Python.

Although all ML models' parameters can be fine-tuned to optimize performance and run-time, the default parameters will be used since it is not the aim of the study to create an optimal implementation. The focus is instead on the feature selection's effect on the performance of the classifiers.

3.2 Dataset

The dataset used in this study is a variant of the one produced by the ISIC Archive. The difference with the original one is that the one used only has two classes: Malignant and Benign. The dataset contains around 3400 images, each being 224x224 pixels. The dataset is balanced, which has positive effects on the classifiers accuracy. The larger, but unbalanced, HAM10000 dataset was also considered. However, during initial testing it was discovered that the balanced dataset resulted in a better accuracy. It was also easier to work with the balanced dataset, as it already only had two classes while HAM10000 has 7.

The dataset was downloaded and structured in order to have every image in the same folder. In order to store the ground truth information indicating whether a given image is malignant or benign, a one-dimensional array was created storing boolean values.

3.3   Metrics

3.3.1   Accuracy (Balanced accuracy)

Accuracy is a metric which is widely used and simple to calculate. It is calculated by dividing the total number of correctly classified images with the total number of images. It therefore measures the probability that the model will classify a given random image correctly. källa

Accuracy was the metric used to determine the performance of the classifiers and the effect of the feature selection methods.

3.4   Feature selection

All feature selection methods used in the study originates from the `Scikit-sklearn` library `feature_selection`.

The manner in which the feature selection was performed differs slightly due to the selection method employed. For the SFS, the feature set was initially empty. The method then adds all features, one at a time. For the SBE the feature set at first contained all features, whereupon features were removed one at a time, until the feature set consisted of only one feature. In each iteration, the accuracy and which features constituted the feature set were recorded. This process was repeated five times for every selection method and classifier. Hence, 40 tests were conducted in total.

3.5   Evaluation

The method for evaluating the three different research questions are stated below.

3.5.1   Is there a difference between different models concerning which features are most effective?

Furthermore, since in every iteration of the selection methods, the included features were recorded, the features included in the feature set generating the optimal accuracy was determined. Onwards, this feature set will be referred to as the optimal feature set.

Based on this statistic, the probability that a feature would be included in the optimal feature set, was calculated by dividing the number of times

the feature was included in the optimal feature set with the number of tests performed. The features with the highest probability can therefore be seen as the most important.

3.5.2 Which features and combinations of features are the most important in terms of classification accuracy?

Based on the findings from the previous research question, the results were summarized to give insight on a more general level where the occurrences of each feature in the optimal feature set was divided by the total number of tests performed. This gives an insight into which features are most likely to constitute the optimal feature regardless of the classifier.

3.5.3 Are some feature selection methods more efficient in creating better accuracy than others? Which one uses the fewest features?

Since multiple tests were performed for every combination of classifier and selection method, it becomes possible to calculate the average accuracy for each number of features in the feature space. Hence the highest average accuracy and the expected number of features needed to achieve this accuracy can be determined. To determine which selection model is the most efficient, the difference between the original accuracy and the highest average accuracy using feature selection, was examined.

Chapter 4

Results

Placeholder reference in order to test the bibliography: [14]

4.1 Determining important features for models

4.2 Determining most important features

4.3 Difference between selection methods

And then textcite: Einstein [14] Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Chapter 5

Discussion

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Chapter 6

Conclusions

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Bibliography

[1] Strålsäkerhetsmyndigheten. *Sverige har snart samma nivåer av hud-cancerfall som Australien.* `https://www.stralsakerhetsmyndigheten.se/press/nyheter/2021/sverige-har-snart-samma-nivaer-av-hudcancerfall-som-australien/`. Strålsäkerhetsmyndigheten, June 2021.

[2] American Academy of Dermatology Association. *SKIN CANCER.* `https://www.aad.org/media/stats-skin-cancer`. American Academy of Dermatology Association, Apr. 2022.

[3] Esra Mahsereci Karabulut, Selma Ayşe Özel, and Turgay İbrikçi. "A comparative study on the effect of feature selection on classification accuracy". eng. In: *Procedia technology* 1 (2012), pp. 323–327. ISSN: 2212-0173.

[4] Shivangi Jain, Vandana jagtap, and Nitin Pise. "Computer Aided Melanoma Skin Cancer Detection Using Image Processing". In: *Procedia Computer Science* 48 (2015). International Conference on Computer, Communication and Convergence (ICCC 2015), pp. 735–740. ISSN: 1877-0509. DOI: `https://doi.org/10.1016/j.procs.2015.04.209`. URL: `https://www.sciencedirect.com/science/article/pii/S1877050915007188`.

[5] Cancer Research UK. *Stages and types.* `https://www.cancerresearchuk.org/about-cancer/melanoma/stages-types`. Cancer Research UK, May 2020.

[6] Cancer Research UK. *Survival.* `https://www.cancerresearchuk.org/about-cancer/melanoma/survival`. Cancer Research UK, Mar. 2020.

[7] N. C. F. Codella et al. "Deep learning ensembles for melanoma recognition in dermoscopy images". In: *IBM Journal of Research and De-*

velopment 61.4/5 (2017), 5:1–5:15. DOI: `10.1147/JRD.2017.2708299`.

[8]   K. Das et al. "Machine Learning and Its Application in Skin Cancer". In: *International Journal of Environmental Research and Public Health* 18.24 (2021), p. 13409. DOI: `10.3390/ijerph182413409`. URL: `https://doi.org/10.3390/ijerph182413409`.

[9]   Catarina Barata, M. Emre Celebi, and Jorge S. Marques. "A Survey of Feature Extraction in Dermoscopy Image Analysis of Skin Cancer". In: *IEEE Journal of Biomedical and Health Informatics* 23.3 (2019), pp. 1096–1109. DOI: `10.1109/JBHI.2018.2845939`.

[10] J. Jaworek-Korjakowska and P. Kłeczek. "Automatic Classification of Specific Melanocytic Lesions Using Artificial Intelligence". In: *BioMed Research International* 2016 (2016), p. 8934242. DOI: `10.1155/2016/8934242`. URL: `https://doi.org/10.1155/2016/8934242`.

[11] N. Smaoui and S. Bessassi. "A developed system for melanoma diagnosis". In: *International Journal of Computer Vision and Signal Processing* 3.1 (2013), pp. 10–17.

[12] Reda Kasmi and Karim Mokrani. "Classification of malignant melanoma and benign skin lesions: implementation of automatic ABCD rule". In: *IET Image Processing* 10.6 (2016), pp. 448–455. DOI: `https://doi.org/10.1049/iet-ipr.2015.0385`. eprint: `https://ietresearch.onlinelibrary.wiley.com/doi/pdf/10.1049/iet-ipr.2015.0385`. URL: `https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/iet-ipr.2015.0385`.

[13] Gareth James et al. *An Introduction to Statistical Learning*. New York, NY: Springer, 2013.

[14] Albert Einstein. "On the Relative Value of Relatives". In: *J. Irreproducible Results* 17 (2016), pp. 4711–4721.

Appendix A

Something Extra