UNIVERSITY OF LJUBLJANA
SCHOOL OF ECONOMICS AND BUSINESS

MASTER’S THESIS

# THE USE OF NEURAL NETWORKS AND DEEP LEARNING IN INSURANCE

Ljubljana, March 2021 MAJA FURLAN

# AUTHORSHIP STATEMENT

The undersigned Maja Furlan, a student at the University of Ljubljana, School of Economics and Business, (hereafter: SEB LU), author of this written final work of studies with the title The Use of Neural Networks and Deep Learning in Insurance, prepared under the supervision of Prof. Aleš Popovič, Ph.D., and co-supervision of Prof. Mauro Castelli, Ph.D.,

D E C L A R E

1. this written final work of studies to be based on the results of my own research;
2. the printed form of this written final work of studies to be identical to its electronic form;
3. the text of this written final work of studies to be language-edited and technically in adherence with the SEB LU's Technical Guidelines for Written Works, which means that I cited and / or quoted works and opinions of other authors in this written final work of studies in accordance with the SEB LU's Technical Guidelines for Written Works;
4. to be aware of the fact that plagiarism (in written or graphical form) is a criminal offence and can be prosecuted in accordance with the Criminal Code of the Republic of Slovenia;
5. to be aware of the consequences a proven plagiarism charge based on this written final work could have for my status at the SEB LU in accordance with the relevant SEB LU Rules;
6. to have obtained all the necessary permits to use the data and works of other authors which are (in written or graphical form) referred to in this written final work of studies and to have clearly marked them;
7. to have acted in accordance with ethical principles during the preparation of this written final work of studies and to have, where necessary, obtained the permission of the Ethics Committee;
8. my consent to use the electronic form of this written final work of studies for the detection of content similarity with other written works, using similarity detection software that is connected with the SEB LU Study Information System;
9. to transfer to the University of Ljubljana free of charge, non-exclusively, geographically, and time-wise unlimited the right of saving this written final work of studies in the electronic form, the right of its reproduction, as well as the right of making this written final work of studies available to the public on the World Wide Web via the Repository of the University of Ljubljana;
10. my consent to the publication of my personal data that are included in this written final work of studies and in this declaration, when this written final work of studies is published.

Ljubljana, March 3rd, 2021 Author's signature:

Maja Furlan

# TABLE OF CONTENTS

## LIST OF FIGURES

# LIST OF ABBREVIATIONS

sl. – Slovenian

**API** – (sl. vmesnik za namensko programiranje) application programming interface

**CNN** – (sl. konvolucijsko nevronsko omrežje) convolutional neural network

**DL –** (sl. globoko učenje) deep learning

# INTRODUCTION

Insurance as a means of self-protection has been a part of human lives for a very long time. Organized joint protection against the risks we face represents higher safety for the ones included in the insurance. For that reason, people turn to insurance to improve economic security. Organizations that provide financial compensations in the case of loss arrange contracts that define the elements that are insured and the conditions in which the loss would be compensated. These contracts are known as insurance policies. The policy is concluded with the signatures of both parties, which represent consent of all terms and conditions in the contract (Ivanjko, Ivanjko, Ivanjko, & Ihanec, 1999).

From the beginnings of the insurance organization, technology has made enormous progress. People keep producing new inventions that simplify their work in all areas, including insurance. Taking advantage of functions that computers can perform is one part of the simplification of the work, which is used for a few decades. Due to the constant improvement of technology, new opportunities are regularly emerging. One of these inventions is artificial neural networks. They were inspired by biological neural networks, which are very capable and can learn from the environment during the course of life. Artificial neural networks are constructed as algorithms that can solve many kinds of problems. Unlike ordinary inflexible programs with algorithms for solving only specific problems, artificial neural networks can learn and improve. This way of improving neural networks with training is called deep learning. This method can be used in many fields, including insurance. My master's thesis focuses on the use of neural networks and deep learning for assuring that the insurance policies are correctly concluded.

The purpose of my research is to determine what the biggest problems insurance companies face regarding not signed policies are, to find a solution, which would decrease negative impact of not signed policies, and to recognize the significance of the solution for the company and insurance assistants currently performing the signature check task.

To achieve the purpose of my research, I aim to create a successful deep learning model for recognizing signatures on insurance policies to show that the model would be useful in practice for insurance companies. I will study the subject of deep learning in order to learn what would be the best way to build the model, analyze which of the available pre-trained models would be the best choice for recognition of signatures and develop the model. One of the aims is to analyze how to optimize the model to increase its efficiency in predicting the existence of a signature. Furthermore, I aim to research the challenges and opportunities of the model, to apply the model to the problem the company is facing and research the impact that the model would have on the client signature check task performance itself, as well as to the time and economic efficiency of the company. Lastly, I aim to recognize how the model could further help other processes behind the operation of insurance companies.

Thus, to reach the purpose of my master's thesis, I define 3 research questions:

1. How to improve the process of inspecting insurance policies with the use of deep learning while maintaining the accuracy of performed client signature check?
2. What will be the benefits of automation in terms of time and economic efficiency of the company?
3. What are the shortcomings and risks while putting the model to use?

I examined relevant literature on neural networks, deep learning, and insurance, which gave me insight into the latest findings and analyses in these two fields. With the newly acquired knowledge, I applied my findings to the research. I investigated all relevant information about the company and business process that my model would affect. This way, I tried to find the best option of the deep learning model to create.

The data for my research was taken from insurance company Generali zavarovalnica d. d. From their electronic archive database of insurance policies, I acquired approximately 2000 (anonymous) motor vehicle insurance policies. Using the program InDoc Viewer, I extracted a scanned page on which the signature of the client should be and transformed it into a picture format (such as .tiff or .jpeg). The pictures were divided into two groups: training and test group. Pre-trained models that are available in the open-source neural network library Keras were the basis for my model. Thus, I created a deep learning model that determines whether the insurance policy in question has a client signature or not.

When the model was finalized, I presented the model and my findings to the manager of the business process Operating With Insurance Policies. I explained my discoveries and thoughts on the application of the model to certain business processes. We discussed the benefits and shortcomings of putting the model to use and how this automation would improve the time and economic efficiency of the company. For this discussion, I used a semi-structured interview. The information, obtained from the interview, together with previously acquired theoretical knowledge, were the basis for my elaboration of further opportunities of the model in this business process, as well as other business processes throughout the company. In the end, I combined all gained knowledge and made propositions for further improvement and potential expansion of my contribution to this problem/field.

I start my master's thesis with an introduction to insurance, insurance procedures, and the Obligations code. I define some of the important terms in insurance. Then, I describe the business processes that are needed for concluding the insurance contract and the importance of the client's signature at the end of the insurance policy. I present the problem that arises with the lack of client signature on the policy. I conclude the chapter with an explanation of the significance of the automation of client signature check task, currently performed by insurance assistants at the company.

In the second part, I introduce the subject of neural networks and deep learning. I present how learning and training of the deep learning model are carried out. Then, I introduce the term transfer learning and the database ImageNet. I specify some of the pre-made models that I used for transfer learning within my research. I explain how measurement and optimization of the deep learning model are performed and I conclude by presenting the software, which was used for my research.

In the third chapter of my thesis, I combine the field of neural networks and deep learning with insurance. I explain how the field of insurance could benefit from using deep learning models. I present the challenges and opportunities of such a model using SWOT analysis. Furthermore, I present future possible uses of the deep learning model in insurance.

The fourth part of my thesis focuses on my research. First, I present company Generali zavarovalnica d. d., from where I obtained data for my research and I illustrate the issue with missing client signature on the insurance policy. I demonstrate how I obtained the data needed for my research. Then I present the construction of the deep learning model and how it works. I present the outcome of my research with graphic demonstration and I answer the research questions with the results of the generated deep learning model. Lastly, I illustrate the usability of the model in the company using the internal information from the company that I obtained with the semi-structured interview with the head of the sub-department Control and Modification of Insurance Policies at Generali zavarovalnica d. d.

# 1 INSURANCE

Caring for self-protection has always been a part of human lives. Individuals wanted to take care of themselves and their families. They determined that with organized joint protection higher safety can be achieved, especially if all contribute to a manner of protection against the risks they face – insurance. In today's world, insurance is an economic and social activity that aims to induce economic security. It can also be defined as a means of protection from different situations that cause financial loss (Ivanjko, Ivanjko, Ivanjko, & Ihanec, 1999, p. 10).

Here, I define a few terms:

- Insurer is an organisation or a company that provides and handles insurance.
- Insured (also known as the policyholder) is the person who commits to pay the insurance premium in exchange for financial compensation in the case of loss.

- Insurance policy is a legal agreement between the insurer and the insured where the first is obliged to pay financial compensation in the case of loss determined in the insurance policy in exchange for insurance premium paid by the insured at the time of conclusion of the policy (Šker, 2010, p. 20).
- Insurance premium is the amount paid by the policyholder to the insurance company specified in the insurance policy (Šker, 2010, p. 32).
- The insurer agrees to indemnify the insured in the event of a covered loss. The insurer offers indemnity against clients' losses.
- To take out an insurance policy: to insure.

In Slovenia, insurance can be divided into two main categories, personal and non-life insurance. The first group, personal insurance, includes all insurance that concerns the life and health of an individual. These include accident, health, and life insurance. The second group, non-life insurance, consists of all insurances of movable and immovable property, liability, as well as financial risks (Šker, 2010, p. 49). Within these groups, there are compulsory and voluntary insurances. Compulsory insurance is mandatory by law for all citizens while voluntary insurance is not mandatory and policyholders voluntarily decide whether to insure.

According to the collected premium, non-life insurance predominates on the Slovenian market. The purpose of non-life insurance is to ensure compensation to the insured for damage that occurred to the property in an event stated in the insurance policy. Non-life insurance can be divided in:

- home insurance
- movable property insurance
- motor vehicle insurance
- liability insurance (Ojsteršek, 2005).

Considering the research of my master's thesis being applied on motor vehicle insurance, I will dedicate a few words to the description of this insurance.

Within motor vehicle insurance, we distinguish between compulsory and voluntary insurance. Compulsory includes all types of liability arising from the use of a self-propelled motor vehicle. The insurance covers material (vehicles, objects, and personal items) and non-material (bodily injuries and death) damage caused to others by the vehicle in the event of an accident. However, it does not cover damage to the vehicle of the person causing the accident. The compulsory insurance is mandatory for all vehicles under the Compulsory Transport Insurance Act (Zakon o obveznih zavarovanjih v prometu (ZOZP), Ur. l. RS, št. 93/07, 40/12, 33/16 in 41/17). The insurance must be taken out before the vehicle is registered. Voluntary motor vehicle insurance provides financial security for the person responsible for a traffic accident where damage to their vehicle occurred. Voluntary insurance also covers events beyond control, such as theft, the damage that occurred at the parking lot, acts of arrogance, and other inconveniences (Generali zavarovalnica d. d., n. d.).

## 1.1 Insurance Procedures and Obligations Code

The insurance premium that is determined according to the extent and type of risk is an essential part of the policy. The insurer and the insured come to a consensual agreement regarding the height and the method of calculation of the premium, as well as the time and the manner of payment. Due to the fact that the premium is essential to the policy, not being paid has legal consequences. An insurance policy is a contract and as a legal document, it must be signed by both entities. In some countries, only the signature of the insurer is enough. In Slovenia, however, the insurance policy must be signed also by the insured. For the insurer, the mechanical signature of the company is enough (Ivanjko, Ivanjko, Ivanjko, & Ihanec, 1999, pp. 44-47). Obligations Code determines that the insurance policy is concluded when both entities sign the document (Obligacijski zakonik (OZ), Ur. l. RS, št. 97/07, 64/16 in 20/18; 925. člen). The Obligations Code contains the basic principles and general rules for all obligation relationships: a relationship between two entities – creditor and debtor.

By signing, the policyholder expresses the will to conclude the insurance policy and agrees with the content of the policy. The insured undertakes to pay the insurance premium to the insurance company while the insurer undertakes to indemnify the insured or a third party in an event specified in the insurance policy (OZ; 921. člen). For the insurance company, the importance of the signatures on insurance policies comes through in cases where the court is needed for the resolution of conflicts between the company and the insured. If the policy is not signed by the insured, the court will resolve conflict in favor of the insured in most cases.

Few cases cause conflict between the insurer and the insured, and, if not resolved, this can lead to a court trial. These cases mostly include unpaid insurance premiums and insurance recourse. A recourse claim is a reimbursement claimed by the insurance company from the person who caused the damage or from the insured for the indemnity that it has previously paid to the insured. The most common are recourses from motor liability insurance, which occur due to the loss of insurance rights. The Compulsory Transport Insurance Act stipulates that in the event of loss of insurance rights, an insurance company that has paid indemnity to the injured has the right to claim reimbursement of amounts paid together with interest and costs from the insured or the person responsible for the damage (ZOZP;

7. člen). Cases that cause loss of insurance rights are determined by law or insurance conditions, such as:

- the driver used the vehicle for a purpose other than that specified in the insurance contract,
- the driver did not have a valid driving license of the category to which the vehicle or group of vehicles belongs when learning to drive,
- the driver's driving license has been revoked or suspended or he has been sentenced to a protective measure prohibiting driving a vehicle of a certain type or category or to a protective measure prohibiting the use of a foreign driving license in the territory of the Republic of Slovenia,
- the driver has driven the vehicle under the influence of alcohol, drugs, psychoactive drugs, or other psychoactive substances above the permitted limit,
- the driver caused the damage intentionally,
- the vehicle driven by the driver was not technically impeccable,
- the driver accidentally left the scene without providing his personal and insurance information (Zavarovalnica Triglav, 2015).

When the insured is in such debt to the insurance company, the latter firstly sends warnings about the unpaid insurance premium or a recourse claim to the insured requesting him to pay his obligation. If the insured does not pay the debt within the set deadline, the insurance company usually sends three reminders before further actions. If the warnings or reminders yield no response, the general recovery procedure is invoked. At this time, the two entities can come to an agreement regarding the debt of the insured or the insured can lodge an appeal (Generali zavarovalnica d.d., 2018). If they cannot resolve the conflict and come to an agreement, the procedure gets taken to the court based on reliable documents. The company’s clerk prepares all the necessary documentation for filing a lawsuit and hands it over to the lawyer who handles the company’s legal cases. The lawyer then files the lawsuit in the competent court. In both cases, in the recovery of the premium and the recovery of the recourse claim, an insurance policy signed by the insured is also among the documentation that the clerk must prepare for the lawyer (Sodstvo Republike Slovenije, 2018). If the insurance policy is not signed, the lawsuit ends, ordinarily in favor of the insured, due to the contract not being legally valid leading the insurance company to revenue loss (Rozman, 2017). This negatively affects company financial results and leads to lower profits for shareholders of the company. Therefore, to the insurance company, the need for the contract to be signed is great and the signature check at the conclusion is crucial.

## 1.2 Business Processes and Electronic Archive

In this chapter, I will describe how insurance companies store insurance policies and processes in which client signature check is included. Moreover, I will describe the procedures for handling concluding documents after taking out an insurance policy. These procedures are submission of concluding documentation, preparation of documentation for conversion into digital form, capture in the company's archives and further processes of control of concluded insurance policies, and issuing complaints in the case of irregularities (Rozman & Simončič, 2020, p. 2).

Concluding documentation are the documents used in concluding insurance contracts and documents needed to make changes to insurance contracts. These include policies, notices to the policyholder, definition of needs and requirements, requests for changes, annexes to the document, e.g. photographs, statement of damages that occurred, etc. (Rozman & Simončič, 2020, p. 1). Regardless of the sales channel through which they were concluded, all insurance policies with the accompanying attachments are kept in the electronic archive. The latter has to be a legally valid electronic archive under the Protection of Documentary and Archival Materials and Archives Act (ZVDAGA). "This Act regulates the protection of documentary and archival material, the validity or probative value of such material, the protection of public and private archival material as a cultural monument, access to archival material in archives and conditions for its use, tasks of the public archival service and supervision over the implementation of this Act and regulations issued on its basis" (Zakon o varstvu dokumentarnega in arhivskega gradiva ter arhivih (ZVDAGA), Ur. l. RS, št. 30/06 in 51/14; 1. člen).

Among many business processes, subprocesses, and tasks that are required for concluding an insurance policy, there is a client signature check task that insurance assistants must perform (Simončič, Hrvatin, & Siter, 2017).

When the insurance agent concludes an insurance policy with the client, he is obliged to hand over the policy and the accompanying documentation to the insurance company (Rozman & Simončič, 2020, p. 4). If the policy is signed in the classic way, i.e. in physical form, the agent can submit the concluding documentation in person at the insurance company or by mail. In doing so, he is obliged to tie any attachments to the insurance policy. The insurance assistants examine and edit the received concluding documentation daily. They check and, if necessary, complete all metadata that is important for storing policies in the electronic archive. Then, they submit the concluding documentation for scanning and storing in the electronic archive (Rozman & Simončič, 2020, p. 7).

In the case that the agent used an application that enables digital signature at the time of conclusion and the policy was electronically signed via signature tablets, the submission of documentation is done automatically. After confirming the signed contract, the insurance policy is automatically archived in the electronic archive (Rozman & Simončič, 2020, p. 4).

Therefore, the concluding documentation is converted into digital form in any case before further management of the policy and can be accessed as such for the time that is stated by internal rules made according to the legislation of the Republic of Slovenia (Rozman & Simončič, 2020, p. 2). In the digital form, it is available in a legally compliant electronic archive, which has been confirmed by the Archives of the Republic of Slovenia (Rozman & Simončič, 2020, p. 8).

The procedure of concluding the insurance contract and digitalizing its documentation is followed by a review of the correctness of the concluded insurance policies. Insurance assistants are obliged to check the correctness and completeness of all elements in the policy and the adequacy of supporting documentation. The control is carried out on the premise of:

- the basis for insurance of individual products (conditions, price lists, etc.),
- instructions for the implementation of individual control procedures which are an integral part of each product,
- valid current company rules (Rozman & Simončič, 2020, p. 10).

The policy is incorrect and an issue of the complaint is prepared in case of:

- missing documentation affecting the amount of the premium,
- missing documentation affecting the liability of the insurance company,
- identified content errors in the general, technical or financial part of the policy (Rozman & Simončič, 2020, p. 10).

Insurance assistants inspect every insurance policy that was concluded in a non-standard way. They change the status of each insurance policy in the backend information system depending on the findings of the inspection. If every element in the policy is correct and made following the insurance basis, price lists, instructions for the implementation of individual procedures, and valid rules established by individual areas, the status of the policy is changed to valid. In the case that something in the policy is incorrect, the insurance assistant issues a complaint and changes the status of the policy to a complaint. A complaint is a record of irregularities or inconsistencies at the conclusion or submission of concluding documentation (Rozman & Simončič, 2020, pp. 10-11). The insurance agent that concluded the insurance contract receives the complaint and is obliged to resolve it. In the event of a missing signature, the agent contacts the client and obtains the signature.

Insurance policies concluded by the standard procedure are automatically checked by the information system. The status is changed to valid automatically in the process of confirming the submission of the concluding document and importing the documentation into the electronic archive. The conditions of conclusion and control according to the standard procedure are determined by those services that develop insurance products. If the information system detects a deviation from the criteria for the standard procedure, it marks the policy as non-standard (Rozman & Simončič, 2020, p. 11).

Automation of the client signature check task would be beneficial in both procedures of concluding insurance contracts (non-standard and standard). In the non-standard procedure, it would facilitate the work of insurance assistants and would reduce the possibility of human error. In the standard procedure, however, the automation of the task could be even more useful. This procedure is performed automatically and, thus, has currently no additional client signature control. Therefore, an automated client signature check would improve the efficiency of the process.

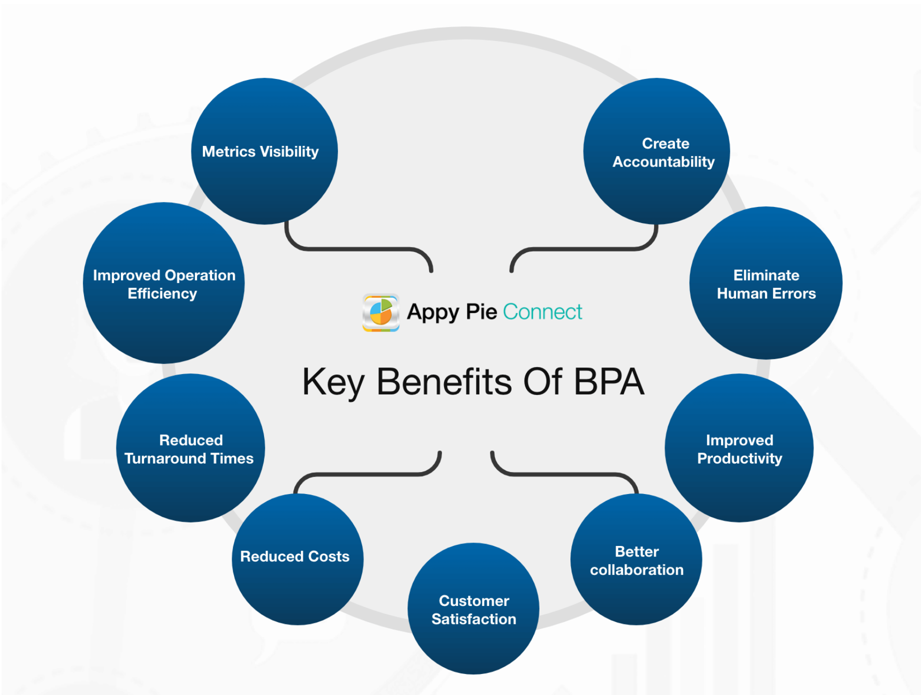
### 1.3 Automation significance

The client signature check task is a routine and easy procedure that must be done on every insurance policy. The task is the same for every policy. The insurance assistant looks at the end of the policy and determines whether it is signed or not. The dullness of the task can lead to monotony and boredom at work. According to Thackray (1981), these two negative side effects can affect performance and quality of work. If the client signature check task would be automated and transferred to be executed by computers, insurance assistants could direct their focus to more demanding tasks that cannot be performed by computers, such as reviewing, if the policy conditions are correct, confirming that the premium height is proportional to risk stated in the policy, finding out if any potential discount is appropriate, etc. Thus, assistants would have more time for these tasks and consequently, the tasks would be done with better quality and fewer errors.

Without automation of the business processes and with reliance solely on human work, both response time and process integrity can suffer. With a higher workload, people are more likely to make mistakes. Automation allows employees to shift their thoughts from administrative tasks to tasks that create added value for the organization (Svet kapitala, 2018).

Automation has many other benefits as can be seen in Figure 1. The most important advantages, concerning automation of client signature check task on insurance policies, are the elimination of human errors, improved productivity, reduced costs, the reduced operating time of the process, and improved operational efficiency. All these perks are significant for the insurance company. For this reason and due to its routine and repetitive nature, the signature check task is an activity suitable for automation (Jakupović, 2018).

*Figure 1: Benefits of Business Process Automation*




*Source: Shukla (2020).*

Insurance policies by one insurance company usually have a standardized form. The logo of the company is always in the same position on the policy. Then, there is the address of the office, the name of the insurance, etc. However, different types of insurances include various diverse things and, therefore, have different structures of the standardized form. For that reason, the deep learning model I created for my master's thesis is taught on only one type of insurance policy. Insurance policies of motor vehicles are only valid for one year. Therefore, there are many concluded each year. For that reason, I chose to use the dataset of insurance policies of motor vehicles for my deep learning model to learn on . In addition to the name of the insurance, the policy includes the policy number, information about the vehicle and the owner, and the dates of the validity of the policy. Then, on the policy, the objects incorporated in the insurance are listed. These include additional elements that the insurance cover, such as insured glass of the vehicle or mirrors, discounts, etc. After these elements, the premium amount, other requirements, and conditions are written in the policy. In the end, there are the signatures. Usually, there is the signature of the insurance agent that concluded the contract on the left side. In the middle, there is the stamp of the company and on the right, there is the signature of the insured. Being that the client signature is always in the same position at the bottom of the policy, insurance policies of motor vehicles are suitable for the deep learning model to learn on. Thus, I aim to automate the routine but necessary task of checking whether the client signed the policy.

# 2 NEURAL NETWORKS AND DEEP LEARNING

In this chapter, I will briefly outline the history of artificial neural networks and introduce the area of neural networks and deep learning. I will describe the ability of learning and training of a deep learning model and introduce the term transfer learning together with an example of a large database that makes transfer learning possible. I will then describe the process of measurement and optimization of results and conclude the chapter with the presentation of the deep learning software I used for creating the deep learning model for the case from the company Generali zavarovalnica d. d.

## 2.1 History of Artificial Neural Networks

Biological neural networks are very complex structures, capable of brilliant problem solving. Scientists play with the idea to transfer the functions of biological neural networks to machines for a very long time. However, the basis for artificial neural networks as we know them today is considered to be the paper "A Logical Calculus of Ideas Immanent in Nervous Activity" by Warren McCulloch and Walter Pitts. They were first to explain how machines could solve problems by imitating mental functions with the use of logic and computation (Piccinini, 2004).

"Artificial neural networks were originally introduced as very simplified models of brain function, so it is initially instructive to consider the broad analogies between artificial neural networks and biological ones." (Bailer-Jones, Gupta, & Singh, 2001, pp. 1-2). Even though artificial neural networks were based on the biological nervous system in our brains, many modern neural network models nowadays bear very little resemblance to biological ones. The term artificial neural network "relatively loose refers to mathematical models which have some kind of distributed architecture, that is, consist of processing nodes with multiple connections" (Bailer-Jones, Gupta, & Singh, 2001, pp. 1-3).

"Artificial neural networks are algorithms, which have been developed to tackle a range of computational problems. These range from modelling brain function to making predictions of time-dependent phenomena to solving hard (NP-complete) problems" (Bailer-Jones, Gupta, & Singh, 2001, p. 1).

## 2.2 Neural Networks and Deep Learning

"A neural network is a massively parallel distributed processor made up of simple processing units that has a natural propensity for storing experiential knowledge and making it available for use. It resembles the brain in two respects:

1. Knowledge is acquired by the network from its environment through the learning process.

2. Interneuron connection strengths, known as synaptic weights, are used to store the acquired knowledge" (Haykin, 2009, p. 2).

A traditional method for the design of neural networks is to use a learning algorithm that modifies the synaptic weights of the network. The ability to learn gives the network a very important property – to generalize, which "refers to the neural network's production of reasonable outputs for inputs not encountered during training (learning)." (Haykin, 2009, p. 2). The problems neural networks usually solve are very complex. Therefore, the approach commonly used is to decompose the problem "into a number of relatively simple tasks, and neural networks are assigned a subset of the tasks that match their inherent capabilities." (Haykin, 2009, p. 2). Neural networks have many other properties and capabilities. One of the most significant for working with neural networks is also adaptivity which refers to neural networks' ability to be retrained due to changing environmental conditions but to still remain stable (Haykin, 2009). This approach to solving problems with neural networks, which has the ability to decompose the problem into smaller parts and the ability to learn from experience, is called deep learning (Goodfellow, Bengio, & Courville, 2016).

There are more types of neural networks, which apply to different kinds of problems. For problems regarding images, the best one is a convolutional neural network (CNN). It contains an input layer, an output layer, and multiple hidden layers in between (Wikipedia, Convolutional neural network, n.d.). "The network moves through the layers calculating the probability of each output" (Wikipedia, Deep learning, n.d.).

## 2.3 Learning and Training

The property of neural networks to generalize is the most important when it comes to learning. Because of this ability, neural networks are capable of solving problems as long as they are similar to the problems encountered at training. The term learning refers to the changes of the components of the neural network, which happen because of the effect of environmental changes. Besides the most common form of learning, changes in connecting weights, learning could be achieved also by developing new connections, deleting existing connections, changing the threshold values of neurons, and more. We can perform any of these learning paradigms by training synaptic weights. This can be achieved by making rules on how to change weights and then connecting these rules into algorithms.

Connecting these algorithms by programming forms a deep learning model (Kriesel, 2007, pp. 59-60).

There are three different learning methods regarding training data that are applied to the model. If the model is given only the input data, the method is called unsupervised learning. This way, it is the most similar to learning of biological neural networks. Due to the fact that the network has to distinguish patterns itself, however, it does not apply to all problems. Reinforcement learning provides the model the input data, and after learning on a sequence, the model receives also the correct results. In supervised learning, the model learns on the input and also correct output data. The network thus changes the weights during the training, which enables the model to efficiently predict results when given new, unknown, but comparable input data. Consequently, generalization makes the latter method the most effective of all three and, therefore, the most used in deep learning (Kriesel, 2007, pp. 61-62). Due to the best efficiency, I will use supervised learning in my model.

In the case, we decide on supervised learning, our data can be organized in a specific way. Input data can be combined into input vector $p = p_1, p_2, \dots, p_n$ called training pattern with corresponding correct output data in vector $t = t_1, t_2, \dots, t_n$. Ordered pairs $(p, t)$ form the set of training patterns $P$. The difference between the output data of the model and the correct output can also be combined into a vector

$$E_p = \begin{pmatrix} t_1 - y_1 \\ \vdots \\ t_n - y_n \end{pmatrix}, \tag{1}$$

called error vector. The objective during training is to minimize the value of components of the error vector as much as possible (Kriesel, 2007, pp. 63-65).

## 2.4 Transfer Learning and ImageNet

For deep learning models to work efficiently, large datasets are needed and accordingly, training takes a long time. Thus, the use of the so-called transfer learning comes in handy. Transfer learning refers to the application of pre-trained structures, i.e. efficient deep architectures, trained with large-scale datasets, to the model, which can be done in few ways. The first uses the fact that “pre-trained filters are adequate to provide sufficient features from a new dataset, and as such, uses pre-trained filters as feature extractor and train the classification part for the new dataset” (Özsert Yiğit & Özyildirim, 2018, p. 351). The second method uses “pre-trained weights as initial values of the learning process” (Özsert Yiğit & Özyildirim, 2018, p. 351). The third option is called fine-tuning. It uses pre-trained weights for some parts of the architecture while “other parts are trained from scratch” (Özsert Yiğit & Özyildirim, 2018, p. 351).

The user chooses the approach of transfer learning regarding the content and size of the dataset. If the user's dataset is small but similar to the pre-trained model, the use of pre-trained filters is favored. In the case the user's dataset is not well comparable to the pre-trained one, the user might be better off training from scratch. In other cases, experts recommend fine-tuning for the first type of dataset from shallow parts and for the least fine-tuning on the whole model (Özsert Yiğit & Özyildirim, 2018, p. 351).

ImageNet is a large image database containing more than 15 million high-resolution images, organized into more than 20,000 categories (Wikipedia, ImageNet, n.d.). ImageNet itself does not store images. Instead, it contains thumbnails and URLs of images (Imagenet, n.d.). Due to the variable resolution of images, they have been downsized to a fixed resolution of 256×256 (Hassan, n.d.).

VGG16 is a convolutional neural network model with 16 layers, constructed by K. Simonyan and A. Zisserman from the University of Oxford. They first presented it in the paper "Very Deep Convolutional Networks for Large-Scale Image Recognition". At the ImageNet Large Scale Visual Recognition Challenge 2014, the model VGG16 achieved 92.7% top-5 test accuracy (Hassan, n.d.). VGG19 is a modification of the VGG model, which consists of 19 layers (OpenGenus, n.d.).

"Inception-ResNet-v2 is a convolutional neural network that is trained on more than a million images from the ImageNet database. The network is 164 layers deep and can classify images into 1000 object categories, such as keyboard, mouse, pencil, and many animals" (MathWork, n.d.).

"Generally, networks trained on ImageNet datasets for ImageNet Large-Scale Visual Recognition Challenges are utilized as pre-trained models" (Özsert Yiğit & Özyildirim, 2018).

### 2.5 Measurement and Optimization

There are a few problems the model can encounter during training. On one hand, the model could only memorize outputs. This means that the model can yield exactly the right results for the data that was used during training but could not effectively predict the right outputs for other data of the same class (Kriesel, 2007, p. 65). The other problem we can encounter during training is overfitting. According to Oxford Dictionary, overfitting in statistics refers to "the production of an analysis which corresponds too closely or exactly to a particular set of data and may therefore fail to fit additional data or predict future observations reliably" (Lexico, n.d.).

The first and the most used solution is to split the dataset into two subsets: a training set and a verification set. The data from the first subset would be used for the model to learn on while the data from the second subset would test the progress of the model during training. To monitor what happens during the training of the model, it is useful to create a learning curve which can indicate whether the network progresses or not. This can be achieved by calculating different errors, such as Euclidian distance, root mean square, or total error and then monitoring how the error decreases during training. The perfect learning curve should look like a negative exponential function (Kriesel, 2007, pp. 65-71).

To know whether the model learns while generating I will measure accuracy and loss. "Accuracy is the count of predictions where the predicted value is equal to the true value." It is expressed in percentage and displays the model's performance. "Unlike accuracy, loss is not a percentage — it is a summation of the errors made for each sample in training or validation sets." Loss is used to calculate weights in the neural network. Within the training process, the objective is to bring the value of loss as close to 0 as possible (Banys & Kobran, n.d.).

## 2.6 Deep Learning Software

### 2.6.1 Python

Python is an open-source, powerful, and widespread programming language (Python, n.d.). It was created by Guido van Rossum and then released in 1991. The language is structured, object-oriented, dynamically typed, and was so designed for writing clear and logical code. It also includes a comprehensive standard library (Wikipedia, Python (programming language), n.d.).

Project Jupyter is a non-profit, open-source project developed in 2014 on GitHub. It supports interactive data science and scientific computing (Jupyter, n.d.). "Jupyter Notebook is a web-based interactive computational environment for creating Jupyter notebook documents" which supports many programming languages, including Python (Wikipedia, Project Jupyter, n.d.).

### 2.6.2 Keras

Keras is a deep learning application programming interface (API) also related to an open-source neural network library written in Python that runs on top of TensorFlow. TensorFlow is an open-source machine learning platform. Together they form "an approachable, highly-productive interface for solving machine learning problems, with a focus on modern deep learning" (Keras, n.d.).

"The core data structures of Keras are layers and models." (Keras, n.d.). Keras offers two ways of building models. The first one is to create the Sequential model (a linear stack of layers). The second one is the Keras functional API. The latter is used to design more complex architectures, such as arbitrary graphs of layers, or write models entirely from scratch (Keras, n.d.).

# 3 USE OF DEEP LEARNING IN INSURANCE

Insurance is an area where diverse opportunities for neural networks and deep learning emerge. There are many complex assignments employees perform that could be assigned to deep learning models. These models would have to be complex as well due to difficult tasks they would take over from employees. The models would have to be made by multiple experts in various fields who would collaborate on this construction. Besides, the immense amount of data would have to be taken into account. To teach the deep learning model how to perform a complicated task so accurately that there would be no need for the revision of employees, the dataset for the learning process would have to include extensive information. However, there are also numerous routine, simple procedures in the operation of insurance companies that would benefit from deep learning models. They would take away some of the workloads of employees and perform monotonous tasks without the employees' help. Thus, a person whose workload was cleared of routine tasks could put their effort into more substantial tasks. These employees would consequently feel more accomplished and important for the company.

The deep learning model, created for my master's thesis, would automate client signature checks on the insurance policies of motor vehicles and thus remove this task from the workload of insurance assistants. This idea promises many benefits for the company and employees currently performing the task. However, before the implementation of the model into the company, it is useful to contemplate all advantages and disadvantages of the application of the model to the operation of the company. Therefore, this chapter will be dedicated to the exploration of the challenges, opportunities, and further uses of the deep learning model in insurance.

## 3.1 Challenges and Opportunities of Deep Learning Model

For more thorough insight into the challenges and opportunities of the use of the deep learning model in insurance, I implemented a SWOT analysis. I decided for using this method because the process of preparing SWOT analysis urges structured and analytical thinking about the product made and gives guidelines on how to analyze whether this product is useful.

In this chapter, I will present my views combined with internal information from the head of the sub-department Control and Modification of Insurance Policies at Generali zavarovalnica d. d. on strengths, weaknesses, opportunities, and threats of the deep learning model I created.

*Table 1: SWOT Matrix of the Deep Learning Model*

| | |
|---|---|
| S: • Speed<br>• Reliability<br>• Fewer errors and risks<br>• Lower operating costs<br>• Higher work ethics | W: • Dependence on technology<br>• Decision cannot be explained<br>• Immense amount of data for learning |
| O: • Use of the model for all forms with handwritten signatures<br>• Higher efficiency of employees | T: • Possibility for becoming obsolete<br>• Cyber attacks and viruses<br>• Software or hardware errors could interrupt the work of the model |

*Source: Own work.*

### 3.1.1 Strengths

- Speed

As previously stated, the client signature check task is a routine, monotonous task. It is performed as the last task of the process of element control of the insurance policy. The client signature check on one policy itself does not take a long time. However, when insurance assistants examine a large number of policies per day, dismissing even a quick task like this can mean a significant difference in time efficiency. The use of the model automates this repetitive task that was previously performed manually and thus shortens the time of the implementation of the insurance policy element control process.

- Reliability

Because the client signature check is a quick, routine task that does not require a lot of effort, it can be overlooked for various reasons. It can happen to insurance assistants that they have a lot on their mind and they focus on mentally demanding tasks, such as checking, if all conditions on the policies are right, if everything on the policy is in order, or if there is some mistake and forget to check the signature at the end. To the deep learning model, such a thing could not happen. All scans that are sent to the model will assuredly be checked for the client's signature.

- Fewer errors and risks

The reliability of the model reduces the possibility of errors and risks caused by the human factor for various reasons.

- Lower operating costs

Automation of client signature check task can reduce time spent on one policy, so the insurance assistants can inspect more insurance policies per day. As follows, the cost for one policy would be reduced and, therefore, the overall operating costs would become lower.

- Higher work ethics

Performing routine and monotonous tasks that do not require a lot of effort can lead to a drop in motivation for insurance assistants. They can start to feel unproductive and not useful. If they concentrate on important assignments rather than on dull tasks, the work ethics can improve significantly, and therefore, they perform better.

### 3.1.2 Weaknesses

- Dependence on technology

In recent decades, technology has increased. By implementing it deeper into the operation of their business, companies have become eminently efficient but also highly technologically reliant. The deep learning model itself needs technology to exist. Thus, implementing it into the company to perform routine tasks would mean even more dependence on technology.

- The decision cannot be explained

Deep learning is a kind of solution that represents a revolutionary step forward in technology. DL models can be taught to solve diverse problems in various fields, from science to business and economics. However, the downside of the deep learning model is that it does not provide any explanation of the given results. This way, the user can only see the outcome of the model and cannot clearly know how the model came to this conclusion, which can lead to confusion of the user.

- The immense amount of data for learning

For the model to replace human employees in practice, it would need to become extremely efficient – close to 100%. For this kind of accuracy, it would need an immense amount of data to be learned on. Company Generali zavarovalnica d. d. stores a significant amount of insurance policies. Therefore, this kind of extensive dataset would be possible. However, processing all data to a form that would be suitable for the model would be time-consuming.

### 3.1.3 Opportunities

- Use of the model for all forms with handwritten signatures

The model created for my master's thesis focuses on handwritten client signatures on insurance policies of motor vehicles. The model could be upgraded to recognize the client signatures on every kind of insurance policy that the company provides. In addition, the model could be expanded to recognize client signatures on every type of form with a standard structure and a handwritten signature. An example of such would be damage files of motor vehicles and other insured properties.

- Higher efficiency of employees

With the elimination of the client signature check task, insurance assistants can dedicate their time and effort to tasks that are more important, creative, and require more mental input. As such, insurance development and insurance sales can be given as an example. In the long run, transferring the work of the employees from the client signature check to other tasks may mean revenue growth and in that way, employees can contribute more to the success of the company through their work.

### 3.1.4 Threats

- Possibility for becoming obsolete

In recent years, technology has been advancing faster and faster and the solutions that were revolutionary some years ago are becoming obsolete and are being replaced by new ones. Additionally, with the advancement of e-commerce and digital signature, there is less need for a handwritten signature. Therefore, if the handwritten signature becomes obsolete, the model would become useless by default.

- Cyber attacks and viruses

The use of technology has many benefits in terms of automation processes. With all that, however, digital threats also come. With the use of computers and the internet, there always is the risk of computer viruses. In a world where technology is more and more used, cybercrime is on the rise. An increasing number of people take advantage of the dependence of companies on technology and make cyber attacks on them. The company Generali zavarovalnica d. d. has adopted a strong security policy within which it implements many security measures and follows the innovations in security and adapts to them regularly. Nevertheless, the possibility for cybercrime and intrusion always exists.

- Software or hardware errors could interrupt the work of the model

Technology, like everything else, is not without occasional malfunctions. Whether the problem is a software or hardware error, the malfunction could affect the performance of the model. Software errors could impair the connection between the model and the data. For instance, there could be a problem where the scans would not reach the model so the model would not check the signature on them. Moreover, the connection could interrupt the connection in the other way so that the results of the model's signature check would not arrive to the user. In the case of hardware malfunctions, there could be a problem with the server where the model is located and as a consequence, the operation of the model would be disabled.

## 3.2 Further Possible Uses

As mentioned before, firstly, the deep learning model can be expanded to the verification of client signatures on every insurance policy that the company Generali zavarovalnica d. d. provides. Further, it can be upgraded to the verification of client signatures on every form that has a standard structure. Being that these forms usually have the client signature at the same position – at the end of the policy, on the line, specifically intended for client signature – the model could be taught to recognize client signatures on such forms also. This includes damage files of insured properties. Moreover, the deep learning model could be improved so that it would recognize whether the identity of the signatory is correct.

The model could be upgraded also in a way of automatically issuing a complaint. Currently, when insurance assistants perform the inspection of the insurance policy and they come across an error or some missing data, they issue a complaint and change the status of the policy to a complaint. If the deep learning model that checks client signature would be embedded in the inspection process, issuing a complaint only for missing signature would be an unnecessary waste of time. Insurance assistants could use this time for some more demanding work. The deep learning model itself could be upgraded so that whenever it comes across an unsigned policy, it changes the status of the policy and issues a complaint.

Additionally, the deep learning model could be further expanded to some of the other processes in the company. One of the important possibilities for further uses of the DL model that emerges is broadening the model to different areas of business. It could be learned to recognize damage on photos of insured property, such as cars, real estate, and other insured possessions.

With the fast-evolving technology and digitalization of the business world, the field of deep learning in insurance is still quite new. Some insurance companies have already made use of it, even though for rather small tasks, such as the client signature check task. However, the field is slowly expanding and developing by businesses allowing deep learning to automate more tasks. With the further evolution of neural networks, deep learning, and artificial intelligence in general, countless more opportunities will emerge.

# 4 USE OF DEEP LEARNING IN THE CASE FROM GENERALI ZAVAROVALNICA D. D.

In this chapter, I will present the company Generali zavarovalnica d. d. and the problem it is facing with the unsigned policies and client signature check task. I will describe my research of the use of the deep learning model in insurance, explain how I obtained the data needed for the model, and how I created the model. The data I gathered are scans of insurance policies for motor vehicles. For the model to learn better on a small amount of data, I chose to provide the model with only one type of insurance policy. Provided with the same pattern of form structure, the deep learning model can learn better and faster. In this chapter, I will also analyze the results the model yielded to answer the first research question I raised in the introduction of my master's thesis. In the end of this chapter, I will present my findings I obtained during the interview with the head of the sub-department Control and Modification of Insurance Policies at Generali zavarovalnica d. d., regarding the second and third research question.

## 4.1 Description of the Company Generali zavarovalnica d. d.

The company Generali zavarovalnica d. d. is a part of Generali Group which is one of the largest insurance companies in the world and the leading insurance company in Europe (Generali zavarovalnica d.d., O nas: Generali zavarovalnica, 2020). In Slovenia, Generali has been operating for more than 22 years. It offers a wide range of insurances: life, pension, health, accident, and non-life insurances, such as insurance for motor vehicles, real estate, and mobile property. In January 2020, Generali merged with Slovenian insurance company Adriatic Slovenica (Generali zavarovalnica d.d., Zaupaj v svojo pot - predstavitev zavarovalnice Generali, 2019).

The mission of Generali zavarovalnica d. d. is "to be the first choice by delivering relevant and accessible insurance solutions". They "ensure achievement striving for the highest performance" (Generali zavarovalnica d. d., Vizija, poslanstvo in vrednote, n.d.). To fulfill this mission, the company aims for continuous improvement, not only in the area of services they offer to clients but also in the operating of the company itself. Accordingly, they strive to regularly better their business processes, and with advancing technology also optimize and automate tasks and processes, which would increase the efficiency of their employees' work.

The insurance company Generali zavarovalnica d. d. has always been technologically advanced and open to the introduction of new information technologies. Its leaders are aware that the business is constantly developing and the technology is increasingly embedded into the operation of the companies, as well as the customers are increasingly skilled in using technological solutions and expect a modern approach to offering insurance. Therefore, Generali aims to achieve timely informatization, process automation, and digitalization of business to remain competitive, cost-effective, and closer to customers.

With digitalization, the insurance company opened new communication channels for its customers enabling them to conclude policies faster and register damages, pay premiums and get indemnity easier. The insurance company supports innovations and the use of advanced technology, such as artificial intelligence, data analysis, and more. Therefore, the use of deep learning to improve some processes in the company is very desirable.

## 4.2 The Case From Generali zavarovalnica d. d.

As I explained previously, a client signature is a necessity for concluding any kind of insurance policy. When signing, the client acknowledges what the insurance incorporates and what it does not, indicates his understanding of the conditions of the policy, and agrees to the premium amount and the time and method of payment. It is expected of the client to respect this agreement and adhere to the terms of the policy. If there has been a breach of contract agreement by the client, the company is forced to take measures. As previously noted, the cases that invoke conflicts between the client and the insurance company revolve mostly around unpaid insurance premiums and insurance recourse.

In the event the conflict cannot be resolved without legal help, a court trial is necessary. There, all documentation of the concerned insurance is reviewed as a basis for the later verdict. If the policy is not signed, the client may claim he did not agree to the terms of the contract. In such a case, the court would be likely to favor the client due to the missing evidence of his agreement to the contract. Therefore, the insurance company must obtain the client's signature on every concluded policy.

Due to the importance of the client's signature on the insurance policy, every policy that goes through inspection is also reviewed for potential missing signature. During the examination of other elements on the insurance policy, insurance assistants must inspect if the client signed the policy. Client signature check is a routine task that takes only a few seconds per policy. However, considering that around 1500 insurance policies must be reviewed every day, tasks that take only seconds add up to few hours per day. In the case the policy is not signed, the issue of a complaint takes approximately 4 minutes. Although that alone does not seem much time-consuming, it would be beneficial to the company to automate this mundane task.

The client signature check task is performed by insurance assistants that have many other assignments too. Their other duties can be more challenging so the client signature check task can sometimes be overlooked. Without reviewing the client's signature they could dedicate this few minutes or hours to one of the more important assignments they face. This way the company would make better use of every insurance assistant not concentrating on such routine tasks and, at the same time, eliminate the possibility of human error at this task.

In the case of insurance policies concluded by the standard procedure, insurance assistants do not additionally perform client signature checks because the policy is automatically marked as concluded. In the event the client forgot to sign the document, the policy remains unsigned. In such a case the advantage of an automatic client signature check would be great because there would be no more unsigned policies concluded through standard procedures.

The company's management acknowledges that the client signature check task as it is a routine and simple procedure would be a necessary task to optimize. It would encourage higher productivity of employees, lower operating costs, and fewer errors occurring in the process. For that reason, the management of the company is searching for solutions that would somehow automate this process and help improve the efficiency of the process. In my master’s thesis, I propose the solution of using neural networks and the deep learning model to automate the client signature check task.

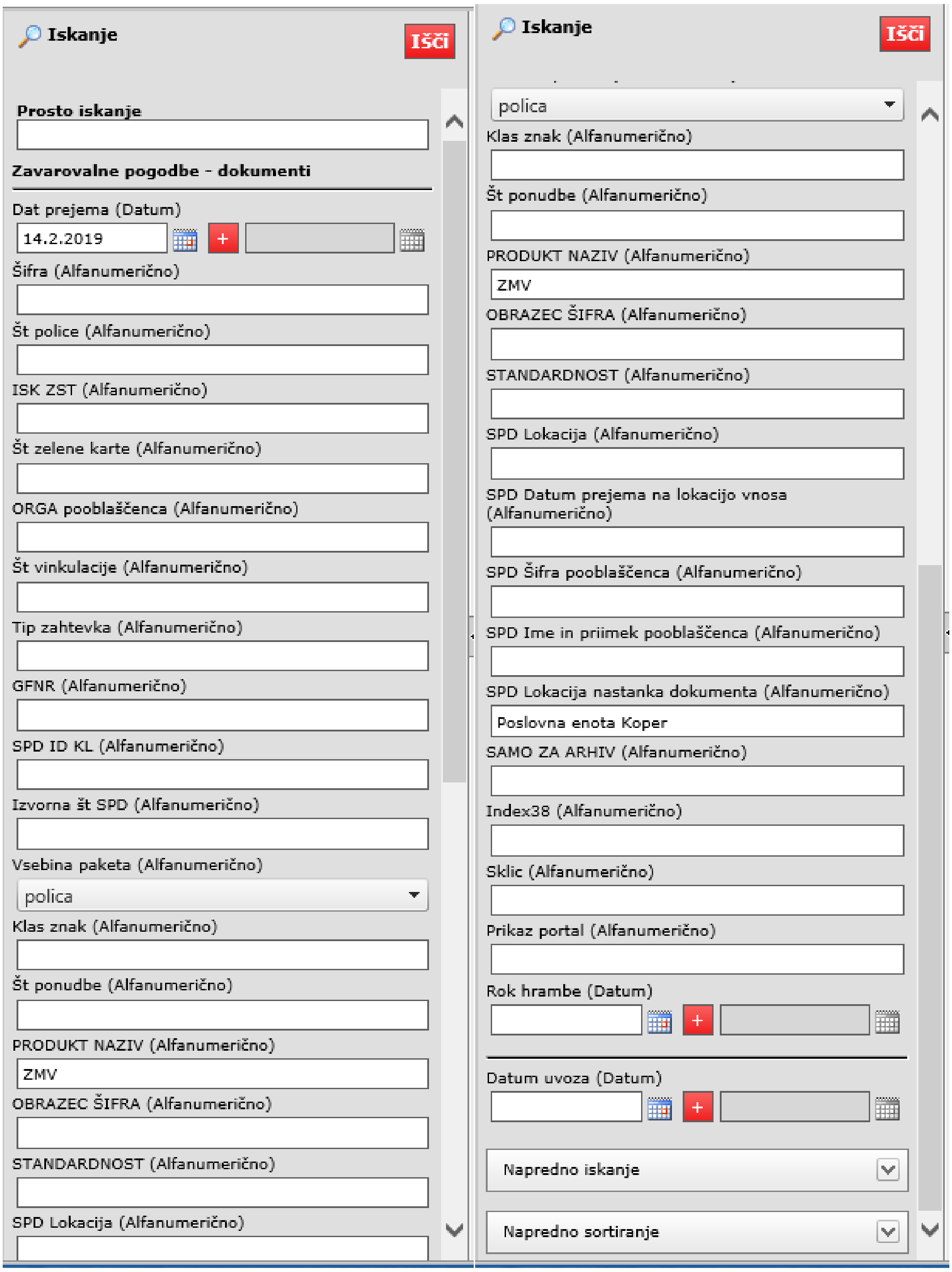
## 4.3 The Model

In this chapter, I will present the deep learning model I created as the purpose of my master’s thesis. I will describe how I obtained insurance policies needed for building the model and how I processed them to be suitable to use during the training of the model. Then, I will describe how I built the model, and at the end of the chapter, I will show two graphs to illustrate the results of the model and provide the results of testing the model on 10 previously unseen scans of insurance policies.

As previously stated, Generali keeps all insurance policies in an electronic archive. I accessed the motor vehicles insurance policies through the program for accessing and reviewing insurance policies – InDoc Viewer. To develop the model I needed only the last page of the insurance policy where the client signature is. In case the policy only has one page, the client signature is on the first page. First, I aimed to obtain signed policies. The program InDoc Viewer offers the possibility to search through all policies by the means of search criteria.

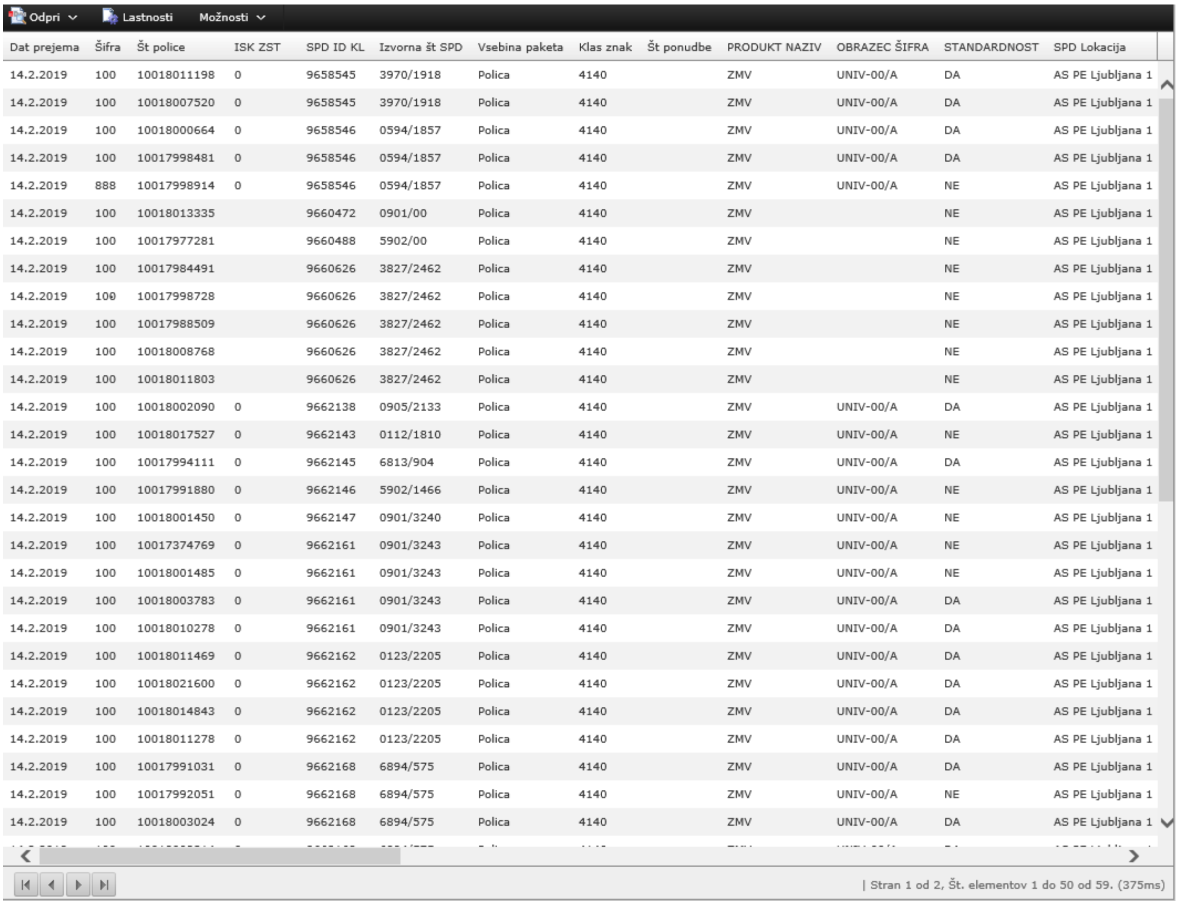
To acquire the signed policies, I selected a random date and a random business unit of Generali. From the policies that the search provided, I extracted the pages with the signature from around 50 random policies.

*Figure 2: Criteria for the Policy Search*




*Source: Mikrocop (n.d.).*

*Figure 3: Result of the Policy Search*




| Dat prejema | Šifra | Št police | ISK ZST | SPD ID KL | Izvorna št SPD | Vsebina paketa | Klas znak | Št ponudbe | PRODUKT NAZIV | OBRAZEC ŠIFRA | STANDARDNOST | SPD Lokacija |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 14.2.2019 | 100 | 10018011198 | 0 | 9658545 | 3970/1918 | Polica | 4140 | | ZMV | UNIV-00/A | DA | AS PE Ljubljana 1 |
| 14.2.2019 | 100 | 10018007520 | 0 | 9658545 | 3970/1918 | Polica | 4140 | | ZMV | UNIV-00/A | DA | AS PE Ljubljana 1 |
| 14.2.2019 | 100 | 10018000664 | 0 | 9658546 | 0594/1857 | Polica | 4140 | | ZMV | UNIV-00/A | DA | AS PE Ljubljana 1 |
| 14.2.2019 | 100 | 10017998481 | 0 | 9658546 | 0594/1857 | Polica | 4140 | | ZMV | UNIV-00/A | DA | AS PE Ljubljana 1 |
| 14.2.2019 | 888 | 10017998914 | 0 | 9658546 | 0594/1857 | Polica | 4140 | | ZMV | UNIV-00/A | NE | AS PE Ljubljana 1 |
| 14.2.2019 | 100 | 10018013335 | | 9660472 | 0901/00 | Polica | 4140 | | ZMV | | NE | AS PE Ljubljana 1 |
| 14.2.2019 | 100 | 10017977281 | | 9660488 | 5902/00 | Polica | 4140 | | ZMV | | NE | AS PE Ljubljana 1 |
| 14.2.2019 | 100 | 10017984491 | | 9660626 | 3827/2462 | Polica | 4140 | | ZMV | | NE | AS PE Ljubljana 1 |
| 14.2.2019 | 100 | 10017998728 | | 9660626 | 3827/2462 | Polica | 4140 | | ZMV | | NE | AS PE Ljubljana 1 |
| 14.2.2019 | 100 | 10017988509 | | 9660626 | 3827/2462 | Polica | 4140 | | ZMV | | NE | AS PE Ljubljana 1 |
| 14.2.2019 | 100 | 10018008768 | | 9660626 | 3827/2462 | Polica | 4140 | | ZMV | | NE | AS PE Ljubljana 1 |
| 14.2.2019 | 100 | 10018011803 | | 9660626 | 3827/2462 | Polica | 4140 | | ZMV | | NE | AS PE Ljubljana 1 |
| 14.2.2019 | 100 | 10018002090 | 0 | 9662138 | 0905/2133 | Polica | 4140 | | ZMV | UNIV-00/A | DA | AS PE Ljubljana 1 |
| 14.2.2019 | 100 | 10018017527 | 0 | 9662143 | 0112/1810 | Polica | 4140 | | ZMV | UNIV-00/A | NE | AS PE Ljubljana 1 |
| 14.2.2019 | 100 | 10017994111 | 0 | 9662145 | 6813/904 | Polica | 4140 | | ZMV | UNIV-00/A | DA | AS PE Ljubljana 1 |
| 14.2.2019 | 100 | 10017991880 | 0 | 9662146 | 5902/1466 | Polica | 4140 | | ZMV | UNIV-00/A | NE | AS PE Ljubljana 1 |
| 14.2.2019 | 100 | 10018001450 | 0 | 9662147 | 0901/3240 | Polica | 4140 | | ZMV | UNIV-00/A | NE | AS PE Ljubljana 1 |
| 14.2.2019 | 100 | 10017374769 | 0 | 9662161 | 0901/3243 | Polica | 4140 | | ZMV | UNIV-00/A | NE | AS PE Ljubljana 1 |
| 14.2.2019 | 100 | 10018001485 | 0 | 9662161 | 0901/3243 | Polica | 4140 | | ZMV | UNIV-00/A | NE | AS PE Ljubljana 1 |
| 14.2.2019 | 100 | 10018003783 | 0 | 9662161 | 0901/3243 | Polica | 4140 | | ZMV | UNIV-00/A | DA | AS PE Ljubljana 1 |
| 14.2.2019 | 100 | 10018010278 | 0 | 9662161 | 0901/3243 | Polica | 4140 | | ZMV | UNIV-00/A | DA | AS PE Ljubljana 1 |
| 14.2.2019 | 100 | 10018011469 | 0 | 9662162 | 0123/2205 | Polica | 4140 | | ZMV | UNIV-00/A | DA | AS PE Ljubljana 1 |
| 14.2.2019 | 100 | 10018021600 | 0 | 9662162 | 0123/2205 | Polica | 4140 | | ZMV | UNIV-00/A | DA | AS PE Ljubljana 1 |
| 14.2.2019 | 100 | 10018014843 | 0 | 9662162 | 0123/2205 | Polica | 4140 | | ZMV | UNIV-00/A | DA | AS PE Ljubljana 1 |
| 14.2.2019 | 100 | 10018011278 | 0 | 9662162 | 0123/2205 | Polica | 4140 | | ZMV | UNIV-00/A | DA | AS PE Ljubljana 1 |
| 14.2.2019 | 100 | 10017991031 | 0 | 9662168 | 6894/575 | Polica | 4140 | | ZMV | UNIV-00/A | DA | AS PE Ljubljana 1 |
| 14.2.2019 | 100 | 10017992051 | 0 | 9662168 | 6894/575 | Polica | 4140 | | ZMV | UNIV-00/A | NE | AS PE Ljubljana 1 |
| 14.2.2019 | 100 | 10018003024 | 0 | 9662168 | 6894/575 | Polica | 4140 | | ZMV | UNIV-00/A | DA | AS PE Ljubljana 1 |

| Stran 1 od 2, Št. elementov 1 do 50 od 59. (375ms)

*Source: Mikrocop (n.d.).*

In this manner, I included 6 different dates and business units from each year between 2016 and 2019. In the selection, I included both standard and non-standard insurance policies. To obtain the policies without a signature, I used an MS Excel file that stores the numbers of the policies, which had the status of complaint (policy without the client signature). I entered each number of the policies with a complaint in the InDoc Viewer and thus obtained a policy page that does not contain the customer's signature. I included all such policies between 2016-2019. In this manner, I extracted 1,030 signed and 1,022 unsigned insurance policies, a total of 2,052 policies.

I saved all pages involved in .tiff format which is provided by InDoc Viewer. Then, I prepared the dataset for processing. I changed the format from .tiff to .jpg for all previously obtained scans of the insurance policy page where the signature should be. I then divided the scans into a learning set and a test set (there are 1920 insurance policies in the learning set and the rest in the test set). I further divided the learning set into two parts, signed and unsigned insurance policies. For this purpose, I wrote a short program code that adds yes or no to the file name. In the programming code for the model, I then further prepared the data for processing in the model.

I made the model in the Jupyter notebook program, which uses the Python programming language. First, I imported all the necessary libraries, packages, and functions for data preparation and drawing graphs.

*Figure 4: Necessary Imports*

```
import numpy as np
import pandas as pd
import os
import cv2
import matplotlib.pyplot as plt
%matplotlib inline
import random
import gc
import matplotlib.image as mpimg
import seaborn as sns
from sklearn.model_selection import train_test_split
```

*Source: Own work.*

Then I wrote a function that takes all the files from the learning set of scans, remembers which ones are signed and which ones are not so that it will learn on this data to distinguish the signature from the blank line for the signature. The program then arranges the data in the learning set in random order. The program also reduces the size of the images as the original size was too large to make the model in a reasonable amount of time. The program then divides the learning set into two more parts on one of which the model will gain knowledge and the other will be used to validate what has been learned.

*Figure 5: Code for Processing Data*

```
train_dir = 'C:\\Users\\Maja\\Documents\\train'
train_ja = ['C:\\Users\\Maja\\Documents\\train\\{}'.format(i) for
i in
           os.listdir(train_dir) if 'ja' in i]
train_ne = ['C:\\Users\\Maja\\Documents\\train\\{}'.format(i) for
i in os.listdir(train_dir) if 'ne' in i]
train_imgs = train_ja + train_ne
print(len(train_ja), len(train_ne), len(train_imgs))
random.shuffle(train_imgs)
```

```
test_dir = 'C:\\Users\\Maja\\Documents\\test'
test_imgs = ['C:\\Users\\Maja\\Documents\\test\\{}'.format(i) for
i in
           os.listdir(test_dir)]
random.shuffle(test_imgs)
del train_ja
del train_ne
gc.collect()

nrows = 200
ncolumns = 250
channels = 3
def read_and_process_image(list_of_images):
    x = [ ]
    y = [ ]
    j = 0
    n = 0
    for image in list_of_images:
        x.append(cv2.resize(cv2.imread(image, cv2.IMREAD_COLOR),
                (nrows,ncolumns), interpolation=cv2.INTER_AREA))
        if 'ne' in image:
            y.append(0)
            n += 1
        elif 'ja' in image:
            y.append(1)
            j += 1
    print(j,n)
    return x, y

x, y = read_and_process_image(train_imgs)
del train_imgs
gc.collect()

x = np.array(x)
y = np.array(y)
x_train, x_val, y_train, y_val = train_test_split(x, y,
test_size=0.20, random_state=2)
del x
del y
gc.collect
```

*Source: Own work.*

For the use of transfer learning, I used the Keras model library. It contains 10 models for classifying images into different categories, which were learned from ImageNet data. The model then adjusts to classify images into two categories: signed and unsigned. For the base pre-trained model I chose to use VGG16. I imported the necessary Keras models and features. With the function ImageDataGenerator I added a transformation to the data – the normalization of the pixel values of the images.

*Figure 6: Imports For Creating the Model*

```
import keras
from keras.applications.vgg16 import VGG16
from keras import Sequential
from keras import layers
from keras.layers import Flatten, Dense, Input
from keras.models import Model
from keras import optimizers
from keras import backend
from keras.preprocessing.image import ImageDataGenerator
from keras.preprocessing.image import img_to_array, load_img

ntrain = len(x_train)
nval = len(x_val)
train_datagen = ImageDataGenerator(rescale=1./255)
val_datagen = ImageDataGenerator(rescale=1./255)
```

*Source: Own work.*

I removed the top layer of the Keras model and froze it to preserve what it had learned on data from Image-net. I added another layer and adapted the model so that it processes a smaller number of images at the same time. The function fit_generator then builds the model and starts the learning process. With the use of the function history, I saved some data for the model. This data includes the accuracy of the model for every epoch and the loss that the model suffered.

*Figure 7: Model Code*

```
batch_size = 32
train_generator_vgg = train_datagen.flow(x_train, y_train,
batch_size=batch_size)
val_generator_vgg      =      val_datagen.flow(x_val,      y_val,
batch_size=batch_size)

conv_base_vgg=VGG16(include_top=False, weights='imagenet',
                                input_shape=(250,200,3))
conv_base_vgg.trainable = False
model_vgg=keras.models.Sequential()
model_vgg.add(conv_base_vgg)
model_vgg.add(Flatten())
model_vgg.add(Dense(256, activation='relu'))
model_vgg.add(Dense(1, activation='sigmoid'))
model_vgg.summary()

model_vgg.compile(loss='binary_crossentropy',
                optimizer=optimizers.Adam(learning_rate=0.002),
                metrics=['acc'])
```

```
history_vgg = model_vgg.fit_generator(train_generator_vgg,
                             steps_per_epoch=ntrain // batch_size,
                             epochs=20,
                             validation_data=val_generator_vgg,
                             validation_steps=nval // batch_size)
```

*Source: Own work.*

*Figure 8: Generating of the Model*

```
Model: "sequential_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
vgg16 (Model)                (None, 7, 6, 512)         14714688
_________________________________________________________________
flatten_1 (Flatten)          (None, 21504)             0
_________________________________________________________________
dense_1 (Dense)              (None, 256)               5505280
_________________________________________________________________
dense_2 (Dense)              (None, 1)                 257
=================================================================
Total params: 20,220,225
Trainable params: 5,505,537
Non-trainable params: 14,714,688
_________________________________________________________________
Epoch 1/20
48/48 [==============================] - 629s 13s/step - loss: 1.4133 - acc: 0.5586 - val_loss: 0.6089 - val_acc: 0.7
500
```

*Source: Own work.*

To understand the model better and how it learned with each epoch of the data, I drew two graphs. One graph demonstrated the accuracy of the model and the other displays loss as they were changing through each epoch during training.

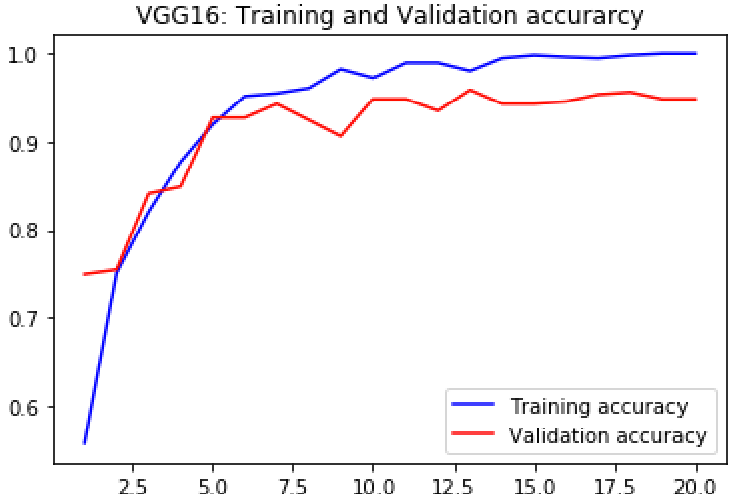
*Figure 9: Code For Graphs*

```
acc_vgg = history_vgg.history['acc']
val_acc_vgg = history_vgg.history['val_acc']
loss_vgg = history_vgg.history['loss']
val_loss_vgg = history_vgg.history['val_loss']

epochs = range(1, len(acc_vgg) + 1)

plt.plot(epochs, acc_vgg, 'b', label='Training accuracy')
plt.plot(epochs, val_acc_vgg, 'r', label='Validation accuracy')
plt.title('VGG16: Training and Validation accurarcy')
plt.legend()

plt.figure()
plt.plot(epochs, loss_vgg, 'b', label='Training loss')
```

```
plt.plot(epochs, val_loss_vgg, 'r', label='Validation loss')
plt.title('VGG16: Training and Validation loss')
plt.legend()

plt.show()
```

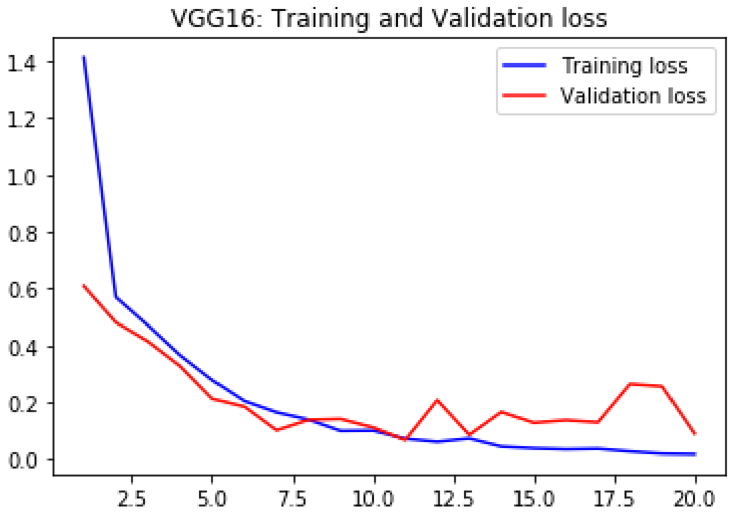
*Source: Own work.*

*Figure 10: Accuracy of the Model*




*Source: Own work.*

*Figure 11: Loss of the Model*




*Source: Own work.*

The values for the set, on which the model was learning, called the training set, are illustrated by the blue line. The values for the validation set are illustrated on the same graph by the red line. The model reached the accuracy of the training set 99.93% and the accuracy of the validation set 94.79% as seen below in Figure 12.

*Figure 12: Results of the Model*

```
Epoch 20/20
48/48 [==============================] - 607s 13s/step - loss: 0.0171
- acc: 0.9993 - val_loss: 0.0890 - val_acc: 0.9479
```

*Source: Own work.*

To see whether the model is equally efficient on previously not seen scans, I used the scans from the testing set.

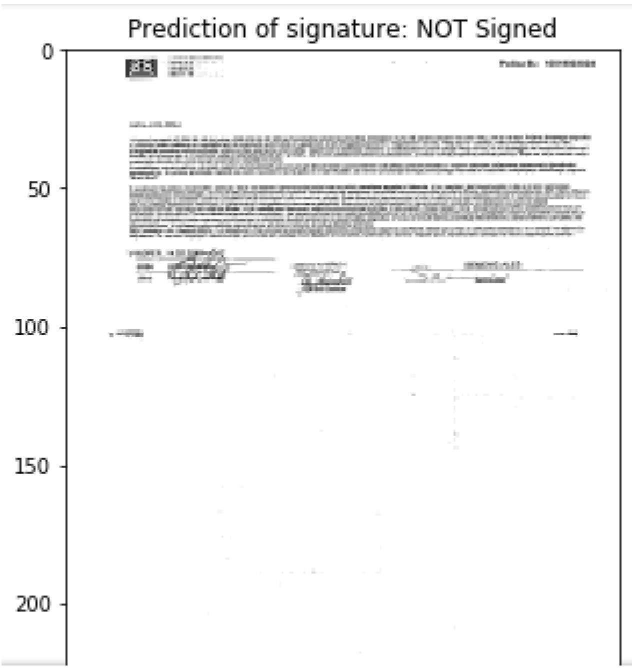
*Figure 13: Code For Testing the Model*

```
x_test, y_test = read_and_process_image(test_imgs[0:10])
x = np.array(x_test)
test_datagen = ImageDataGenerator(rescale=1./255)
i = 0
text_labels = []
plt.figure(figsize=(150,75))
for batch in test_datagen.flow(x, batch_size=1):
    pred = model_vgg.predict(batch)
    if pred > 0.5:
        text_labels.append('Signed')
    else:
        text_labels.append('NOT signed')
    plt.subplot(10, 1, i + 1)
    plt.title('Prediction of signature: ' + text_labels[i])
    imgplot = plt.imshow(batch[0])
    i+=1
    if i%10 == 0:
        break
plt.show()
```

*Source: Own work.*

The program code above tested the model on 10 scans. It predicted the signature on the policy in 9 cases correctly. Therefore, it was 90% accurate. It only yielded one false result seen in Figure 14.

*Figure 14: False Testing Result of the Model*




*Source: Own work.*

## 4.4 Analysis of the Model

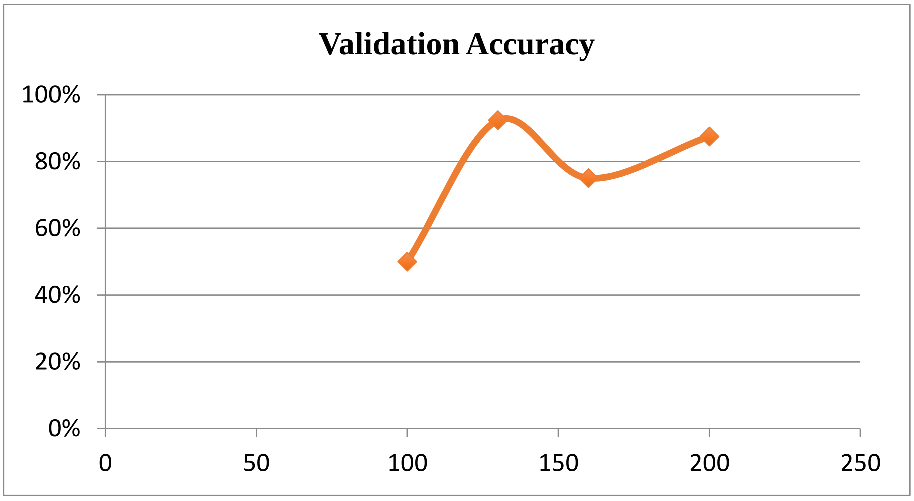
In this chapter I will discuss the findings I attained while trying to answer the first research question that was raised at the beginning of my master's thesis. Because the research question cannot be answered by a simple answer, I divided the evaluation of the model into four parts. I created four questions to help me clarify how to improve the process of inspecting insurance policies with the use of deep learning while maintaining the accuracy of performed client signature check.

- What is the size of the dataset where the model still has a satisfactory success rate and how the success rate of the model increases given more data?

The model generated and learned on 1920 policies. It achieved great results with the accuracy of the validation set 94.79% and the loss of validation set 0.089. However, I was interested in the smallest size of the dataset where the model still presents adequate results – the accuracy of the validation set accuracy 80% or more but also stable accuracy results over each epoch.

I let my model learn consecutively on 100, 130, 160, and 200 scans of insurance policies. Validation accuracy and validation loss of all four datasets can be seen in Figure 15 and Figure 16. The accuracy increased with more scans in the dataset and the loss decreased. However, the escalation of accuracy and diminishment of loss is not properly consistent.

*Figure 15: Validation Accuracy For Different Amount of Data*



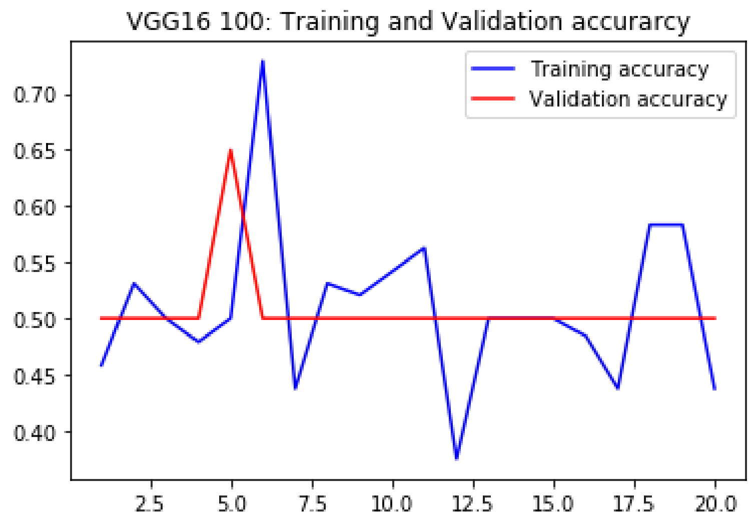

*Source: Own work.*

*Figure 16: Validation Loss For Different Amount of Data*


*Source: Own work.*

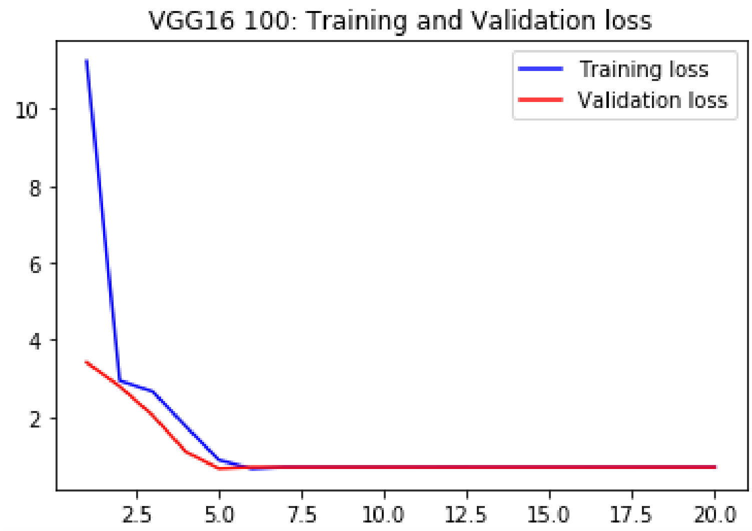
On the dataset with 100 policies, the model did certainly not learn enough to be taken into account. The accuracy of the training set reached only 43.75% while the validation set achieved the accuracy of 50%. The loss of the training set was 0.6938 and the loss of the validation set 0.6932. Both can be seen in Figure 17 and Figure 18.

*Figure 17: Accuracy of the Model With 100 scans*




*Source: Own work.*

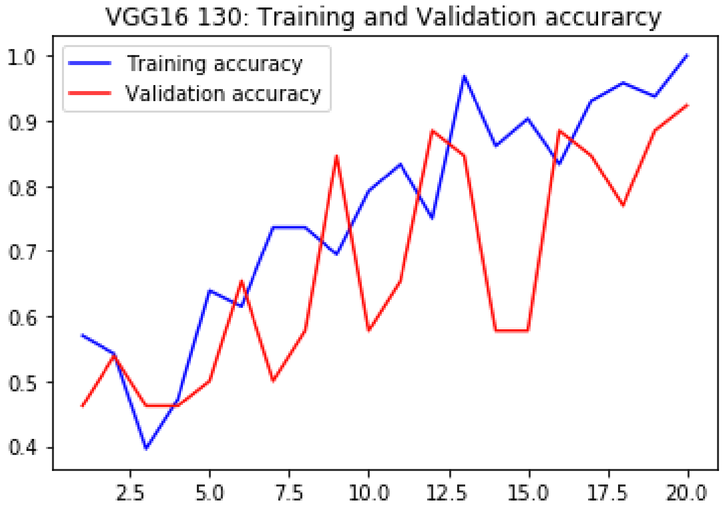
*Figure 18: Loss of the Model With 100 scans*




*Source: Own work.*

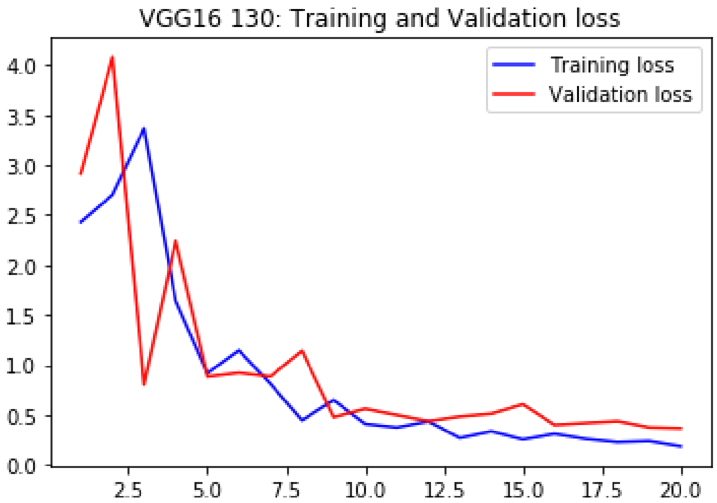
The model taught on 130 policies achieved very high accuracies, training accuracy 100%, and validation accuracy 92.31%. The loss of the training set was 0.1897 and the loss of the validation set was 0.3681. However, the model was very unstable during training at each epoch, seen in Figure 19 and Figure 20. Only two epochs before finishing training the model still had the validation accuracy of 76.92%.

*Figure 19: Accuracy of the Model With 130 scans*




*Source: Own work.*

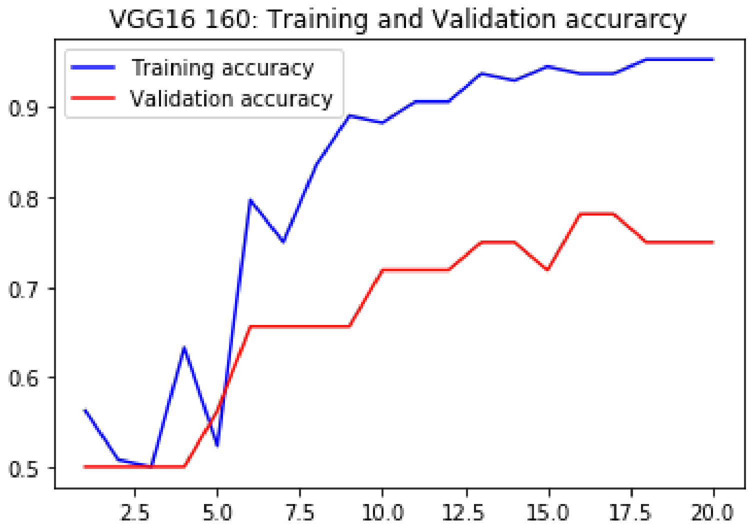
*Figure 20: Loss of the Model With 130 scans*




*Source: Own work.*

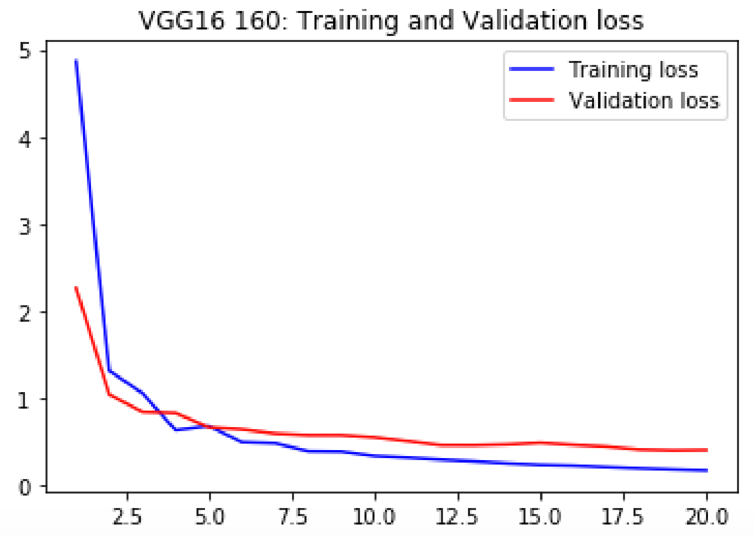
On the dataset with 160 policies, the model reached training accuracy 95.31%, validation accuracy 75%, training loss 0.1778, and validation loss 0.4131 (Figure 21 and Figure 22). Thus, it did not perform well enough to use in practice.

*Figure 21: Accuracy of the Model With 160 scans*




*Source: Own work.*

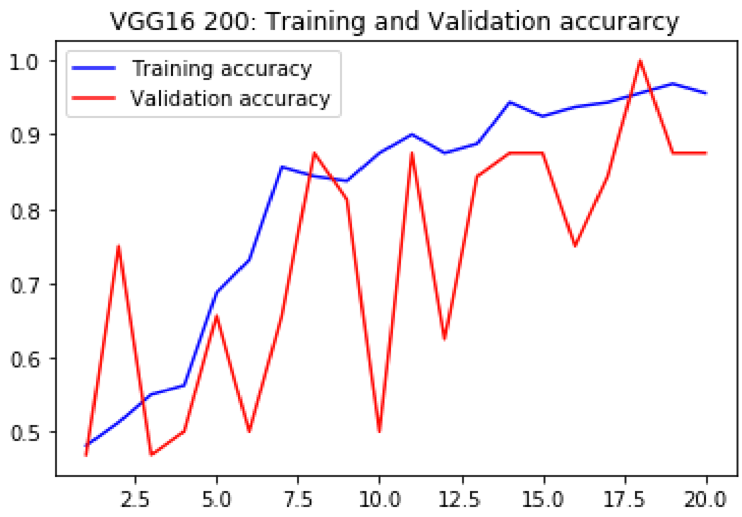
*Figure 22: Loss of the Model With 160 scans*




*Source: Own work.*

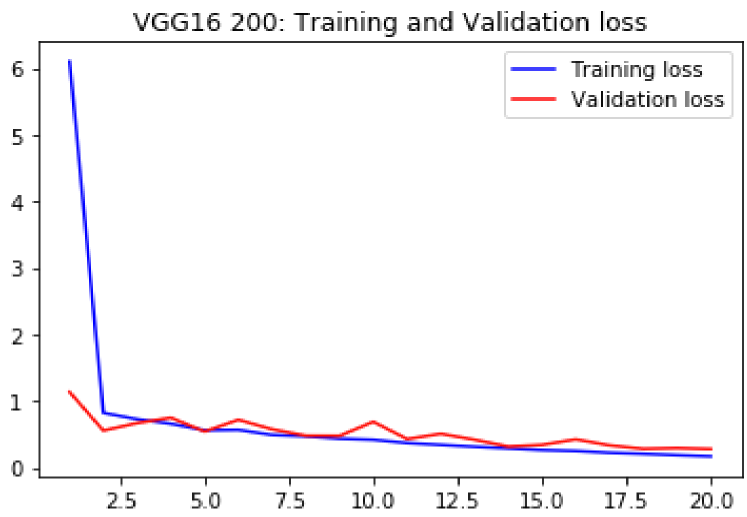
The model that learned on the dataset with 200 policies performed moderately better. It achieved the training accuracy of 95.63%, validation accuracy of 87.50%, training loss of 0.1708, and validation loss of 0.2808. During training, it was slightly more consistent than the models, which learned on smaller datasets (Figure 23 and Figure 24).

*Figure 23: Accuracy of the Model With 200 scans*




*Source: Own work.*

*Figure 24: Loss of the Model With 200 scans*




*Source: Own work.*

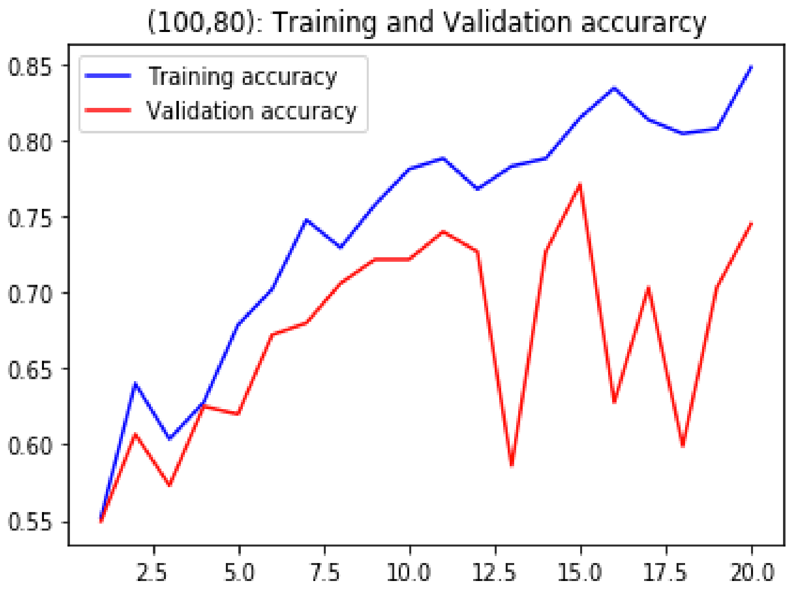
From the models created on smaller datasets, I can conclude that the amount of data is important for the deep learning model to work efficiently. The model with a learning dataset containing almost 2000 scans performed significantly better.

- What is the smallest size of the pictures so that the model is still efficient?

Original scans I obtained had the size (2500,2000). For the model to be generated in a reasonable amount of time and yet achieve a high success rate, the pictures were all downsized to the size of (250,200). The question that was raised was how much can the pictures be downsized but still remain efficient.

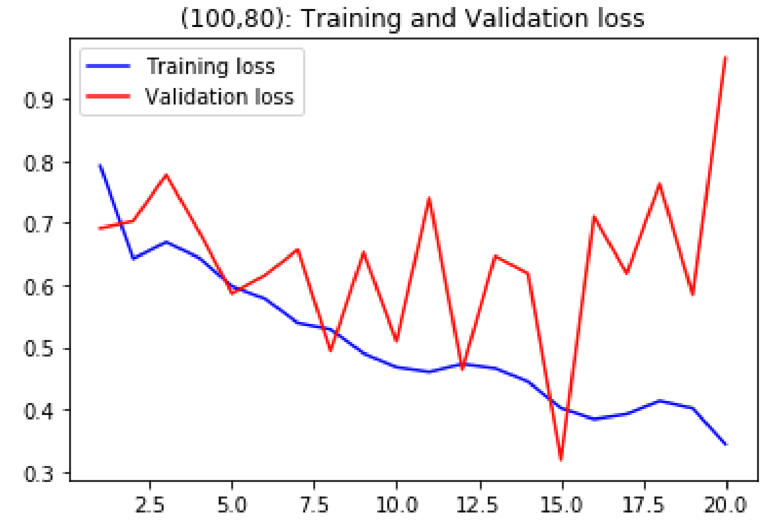
The model that learned on the images sized (100,80) reached training accuracy of 84.77% and training loss of 0.3444. However, it performed badly on the validation set. It reached a validation accuracy of 74.48% and a validation loss of 0.9647. The results can be seen in Figure 25 and Figure 26.

*Figure 25: Accuracy of the Model With the Size (100,80)*




*Source: Own work.*

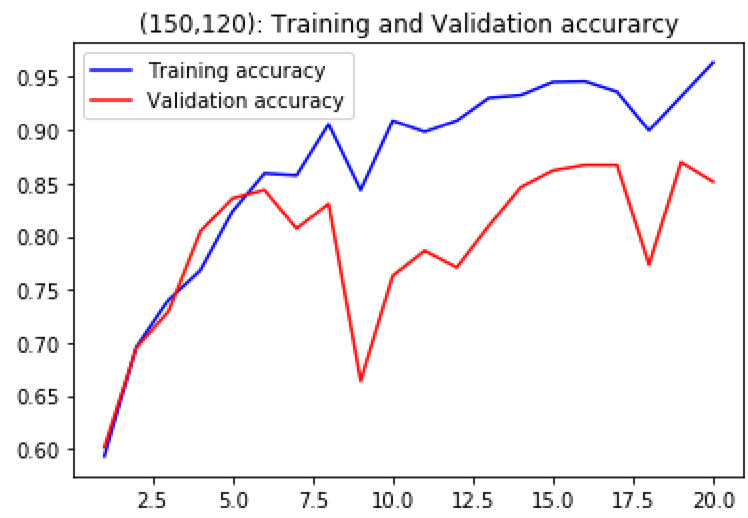
*Figure 26: Loss of the Model With the Size (100,80)*




*Source: Own work.*

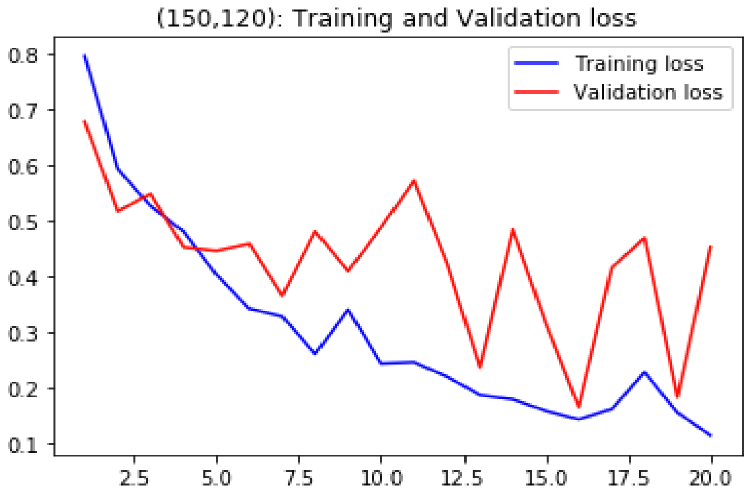
The model that learned on the dataset where the images were sized (150,120), performed better. It achieved training accuracy of 96.35%, validation accuracy of 85.16%, training loss of 0.1145, and validation loss of 0.4525. Yet, it still provided results that were not stable over epochs during training and the validation accuracy and loss were not as good as with the bigger images (Figure 27 and Figure 28).

*Figure 27: Accuracy of the Model With the Size (150,120)*




*Source: Own work.*

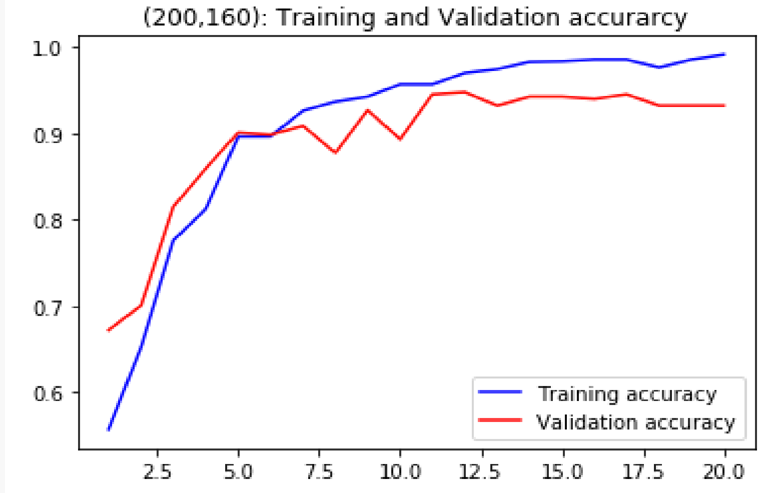
*Figure 28: Loss of the Model With the Size (150,120)*




*Source: Own work.*

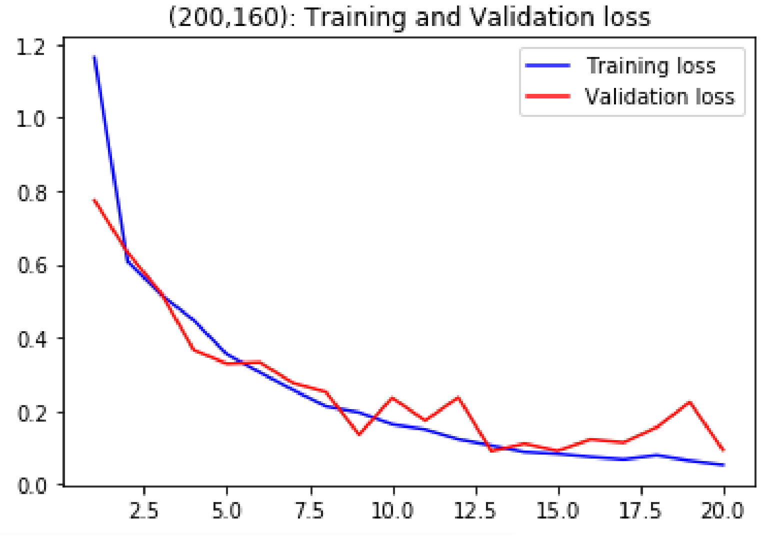
The dataset where the images had the size (200,160) provided adequate results. It reached a training accuracy of 99.15%, validation accuracy of 93.23%, training loss of 0.0529, and validation loss of 0.0936. It achieved results that were close to the model with the size of images (250,200). The results of the model over epochs during training that can be seen in Figure 29 and Figure 30 were relatively stable. Thus, even though slightly bigger pictures still provide better results, the dataset with the image sizes (200,160) would be good enough for the model to use in practice.

*Figure 29: Accuracy of the Model With the Size (200,160)*




*Source: Own work.*

*Figure 30: Loss of the Model With the Size (200,160)*




*Source: Own work.*

- How does the model's stability change with adding layers to the pre-trained convolutional base?

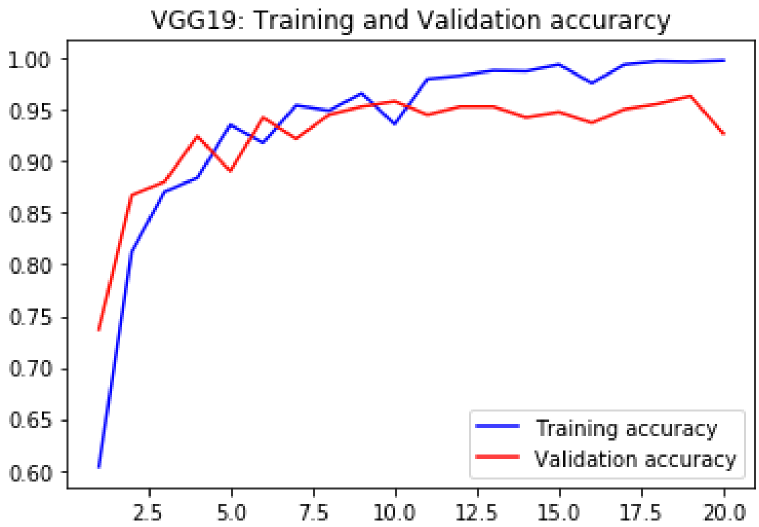
The model did not show significant improvement when added more layers. It performed efficiently. However, the model took longer to generate. Thus, I can conclude that the model I created does not need additional layers to be more efficient.

- Which pre-trained models will be the most successful?

Pre-trained model VGG16 embedded in my model achieved useful results. I wanted to explore whether other pre-trained models perform equally successfully on the type of data that I obtained. For this analysis, I chose to use pre-trained models VGG19, InceptionResNetV2, Xception, and DenseNet121.

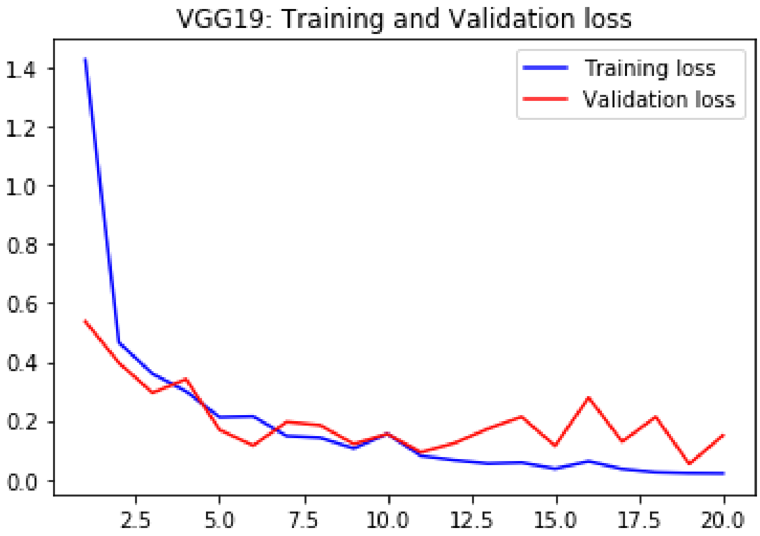
The model with pre-trained model base VGG19 achieved great results. It reached a training accuracy of 99.80%, validation accuracy of 92.71%, training loss of 0.0199, and validation loss of 0.1493 (Figure 31 and Figure 32). However, it did not perform significantly better than the model with the pre-trained model base VGG16.

*Figure 31: Accuracy of the Model With the Base VGG19*




*Source: Own work.*

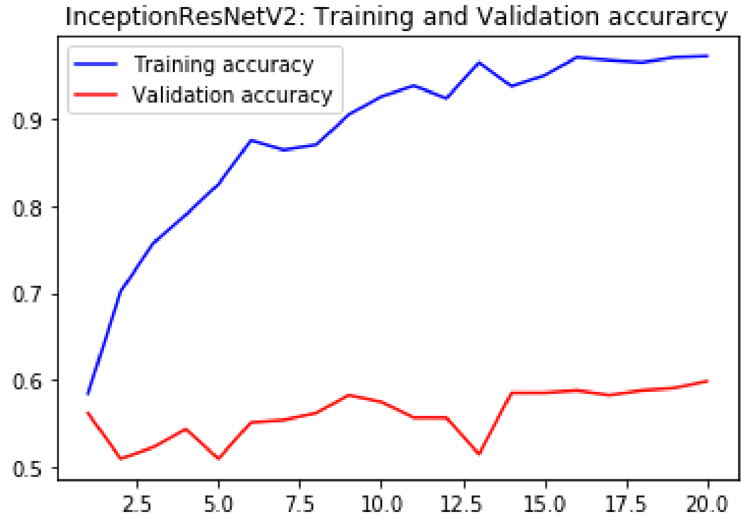
*Figure 32: Loss of the Model With the Base VGG19*




*Source: Own work.*

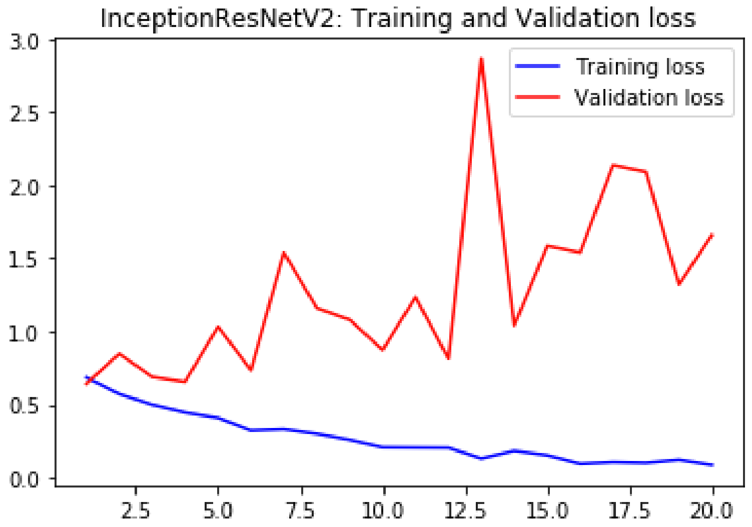
The pre-trained model InceptionResNetV2, however, did not achieve such results. While it learned greatly on the training set, it did not show similar results on the validation set (Figure 33 and Figure 34). The same can be seen for the pre-trained models Xception (Figure 35 and Figure 36) and DenseNet121 (Figure 37 and Figure 38).

*Figure 33: Accuracy of the Model With the Base InceptionResNetV2*



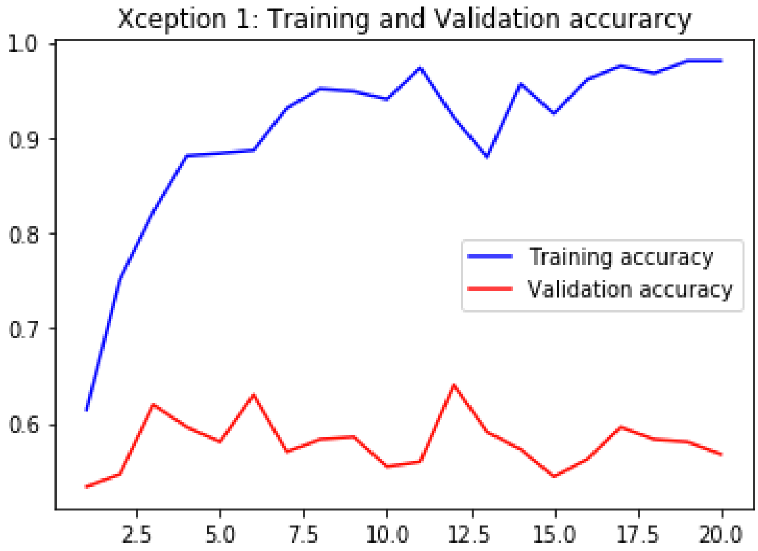

*Source: Own work.*

*Figure 34: Loss of the Model With the Base InceptionResNetV2*



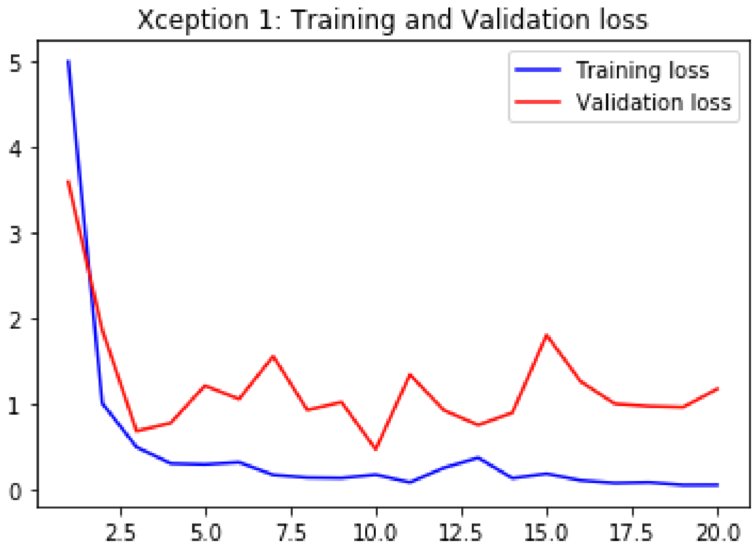

*Source: Own work.*

*Figure 35: Accuracy of the Model With the Base Xception*



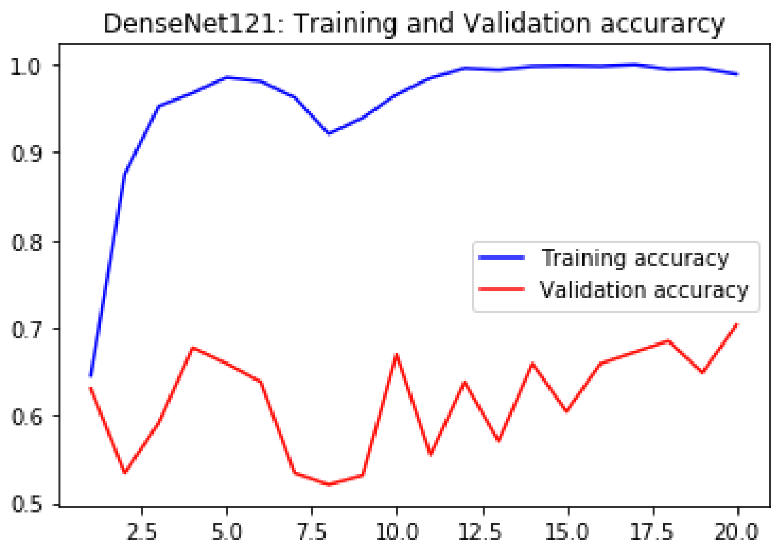

*Source: Own work.*

*Figure 36: Loss of the Model With the Base Xception*



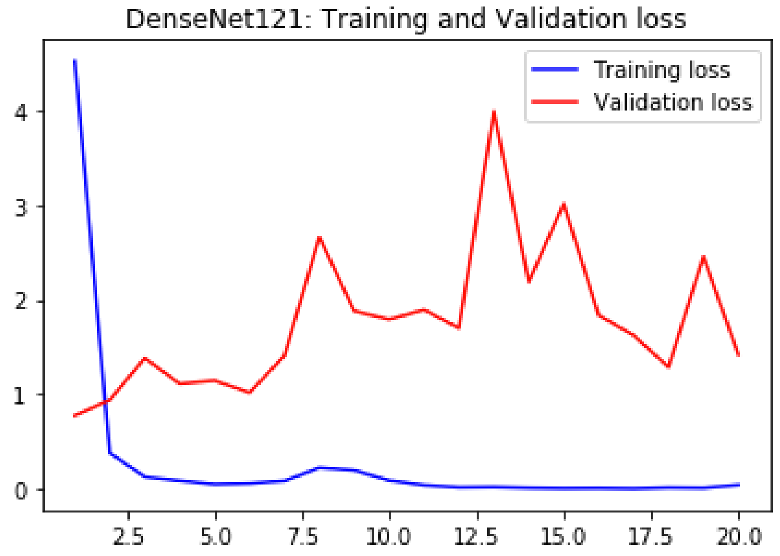

*Source: Own work.*

*Figure 37: Accuracy of the Model With the Base DenseNet121*




*Source: Own work.*

*Figure 38: Loss of the Model With the Base DenseNet121*




*Source: Own work.*

Based on the results of each pre-trained model, I can conclude that pre-trained models InceptionResNetV2, Xception, and DenseNet121 cannot apply to the type of data I obtained. However, models with pre-trained model base VGG16 and VGG19 work excellent and could be used also in practice.

After researching and analyzing the results of the model, I can conclude the answer to my first research question.

1. How to improve the process of inspecting insurance policies with the use of deep learning while maintaining the accuracy of performed client signature check?

Provided with a considerable amount of insurance policies (1,920), the model reaches an accuracy of 94.79%. When training on smaller datasets, the results are less accurate and they improve when provided with a larger dataset. Hence, with an immense amount of scans for the model to learn on, the accuracy could improve even more and thus could reach near-perfect performance. In my research, I concluded, that the scans of insurance policies can have a relatively small size for the model to give accurate results. That means that even if the model would have to check an enormous amount of policies per day, it would still provide results in a considerable amount of time. My research showed that the model does not need extra layers in its neural network to work better. On the opposite, the model with extra layers did not provide significant improvement of results. Instead, it took much longer to generate. For this reason, it is better to use only a few needed layers without additional ones to slow down the generation of the model. However, the model reached high performance only when using some of the pre-trained models. Not all are suitable for the kind of data required to be used in the model. Therefore, the model needs to be built by using either pre-trained base VGG16 or VGG19. The other three pre-trained models did not seem to learn what they should be checking on given scans because they worked great on the training set. Provided with previously unseen scans, however, their performance did not show promising results.

Taking everything into account, I can conclude that with the right pre-trained base for the model and using the right amount of layers, the deep learning model can successfully check the client signatures on insurance policies. The process of inspecting policies remains accurate and becomes faster because the client signature check task can be eliminated from the insurance assistants' workdays. This makes the model a suitable possibility for the insurance company to employ.

## 4.5 Usability of the Model in the Company

For deeper insight into the practicality of the deep learning model that I created, I did a semi-structured interview with the head of the sub-department Control and Modification of Insurance Policies at Generali zavarovalnica d. d. I presented the process of building the model, how it works, and what it does, including the performance and the accuracy of the model. I explained my views on the benefits of implementing my deep learning model into the inspection of insurance policies of motor vehicles. Through the questions of the semi-structured interview, I discovered her observations and opinion of the usefulness of the model. Combining her knowledge and insights with my previous research, I answered the second and third research questions of my master's thesis.

She has been working in the area of insurance for 20 years, more specifically in the field of insurance policy control for over 12 years. At the beginning of her career, the process of managing insurance policies was not done automatically. Insurance agents concluded insurance contracts with clients manually. Insurance assistants then ought to enter each policy into the system separately which ended up as a very time-consuming activity. However, changes are constant in business, especially with rapidly evolving technology. Digitalization and automation have made processes more fluent, as well as more coordinated with other departments and processes in the company.

2. What will be the benefits of automation in terms of time and economic efficiency of the company?

Advances in technology require new skills of employees, as well as new tools for work and new information systems. The department that deals with upgrades is process-oriented and highly specialized. The newly upgraded process tools are very helpful to the department that performs policy control and thus its knowledge gets focused on the content of insurance and the rules of conclusion. The upgraded and new tools make few tasks easier to complete but with the raising competitive business market the workload of employees increases. This requires greater concentration at work and can cause a higher burden on the employees due to the need to know and constantly use a greater load of data.

Verification of the client's signature on the insurance policy is a routine operation when preparing a document for scanning which takes only a few seconds per document. However, employees scan approximately 1,500 documents per day. Checking each one for the signature still takes some time. In the case there is no signature on the policy, it is necessary to issue a complaint about an unsigned document. This requires approximately 4 minutes per case. Therefore, automatic signature check task would result in time savings and consequently to economic savings of the company.

3. What are the shortcomings and risks while putting the model to use?

The only shortcoming or risk the head of sub-department Control and Modification of Insurance Policies at Generali zavarovalnica d. d. could see was the possibility of error in the communication between the model and the user. The users – insurance assistants would have to receive notifications if some insurance policy is not signed. This way, they could meet with the clients and obtain their signatures. If there were no unsigned policies at the end of the day, the model would still have to inform the user. Thus, the assistant would know at the end of each day whether all policies were signed or not. Otherwise, there could be some mistake in the model or the software of the company and the user would not be aware which could lead to larger complications for the company.

# CONCLUSION

Artificial neural networks seem to be an excellent way of solving problems. In the last years, with the expansion of the data on the internet, using neural networks and deep learning is easier than ever. The use of pre-made models, available online, makes it possible for one to create a deep learning model that can solve very complex problems. It can be applied to various fields and insurance is amongst them. Even within the area of insurance, there are diverse opportunities for neural networks and deep learning to simplify the work of employees.

Insurance is a very important area because it provides economic security for people that are insured. It is a means of protection against the situations in which financial loss occurs. For the stable operation of insurance companies, there are many business processes needed. If they want to stay competitive, they need to be consistent with improving their operation as technology advances. Some of the tasks within the processes being automated bring higher time and economic efficiency to the company. The most practical way is to automate routine tasks that are not as demanding as to require much mental effort. Such a task is, among others, a client signature check. The automation of this task would help during the process of concluding insurance contracts. The insurance assistants need to inspect every insurance policy before scanning and one element to inspect is whether the policy is signed. Due to the repetitiveness and routineness of the job, it would be a perfect task to automate and thus eliminate from insurance assistants' assignments. Employees, currently performing this task, would be able to concentrate on more demanding tasks, leaving dull tasks, such as client signature checks, to the machines. This would prevent monotony and boredom at work and cause fewer errors at the client signature check task, additionally leading to the better quality of other performed tasks.

The deep learning model, which would perform the client signature check task, trained on 200 insurance policies, reaches an accuracy of 87.50%. When the amount of the data increases to 1920 policies, the model achieves a very high accuracy of results – 94.79%. The fact that the model improves when presented with more amount of data and that with less than 2000 policies the model provides great results, shows that with a large amount of data, the model could be trained to perform almost impeccable. This makes the model a suitable possibility for the insurance company to employ. The size of the images used can be considerably small for the model to work properly. However, the model reaches such high performance only when using some of the pre-trained models. While models with pre-trained base VGG16 and VGG19 reach results that can easily be implemented in practice, the pre-trained bases InceptionResNetV2, Xception, and DenseNet121 constitute models that are not suitable to use. The models work well on the training set. However, when presented with a validation set, they do not perform satisfactorily. For this reason, the deep learning model I created for my master's thesis should be generated whether with a pre-trained base VGG16 or with VGG19.

In brief, with the right pre-trained base for the model and using the right amount of layers, the deep learning model I created for my master's thesis can successfully check the client signatures on insurance policies. Due to the eliminated client signature check task from insurance assistants' assignments, the process of inspecting policies becomes faster and remains accurate. Therefore, the model is a suitable possibility for the insurance company to employ.

Even though the deep learning model brings some challenges to overcome, such as possibilities for cyber-attacks and viruses, software or hardware errors, and potential overdependence on technology, it offers many significant opportunities to take advantage of. It offers higher speed and reliability of the client signature check task, fewer errors and risks, lower operating costs, and higher work ethics amongst employees. The one thing to be aware of is the communication between the model and the user – insurance assistant. They would have to receive a notification in the event of an unsigned policy, as well as a notification at the end of the day, in case there were no unsigned policies. Otherwise, there could be an error somewhere in the process and the user would not realize it leading to possible complications.

The model I created could be upgraded to recognize signatures also on other insurance policies, to automatically file a complaint about missing signature, or to recognize damage on images of insured properties. For the time being, however, the model without upgrades would benefit the company in the sense of time and economic efficiency. The task that usually takes only a few seconds but can lead to more hours considering the number of policies needed to be inspected every day, can be automated. Thus, the acquired hours can be spent on other, more demanding tasks. This time could be spent more effectively, which means also economic savings for the company consequently.

# REFERENCE LIST


1. Bailer-Jones, C. A., Gupta, R., & Singh, H. P. (2001, 13 February). An introduction to artificial neural networks. *Automated Data Analysis in Astronomy*.

2. Banys, D., & Kobran, D. (n.d.). *Accuracy and Loss*. Retrieved August 31, 2020 from https://docs.paperspace.com/machine-learning/wiki/accuracy-and-loss

3. Generali zavarovalnica d. d. (n.d.). *Avtomobilski kasko*. Retrieved August 31, 2020 from https://www.generali.si/zavarovanje/avto/kasko-zavarovanje-avtomobila#page-top

4. Generali zavarovalnica d. d. (n.d.). *Generali zavarovalnica*. Retrieved August 31, 2020 from https://www.generali.si/aboutus

5. Generali zavarovalnica d. d. (2020). *O nas: Generali zavarovalnica*. Retrieved from https://www.generali.si/generali-zavarovalnica

6. Generali zavarovalnica d. d. (2018, 1 May). *Pravilnik o načinu, pogojih in postopkih izterjave terjatev*. (interno gradivo). Generali zavarovalnica d. d.

7. Generali zavarovalnica d. d. (n.d.). *Vizija, poslanstvo in vrednote*. Retrieved August 28, 2020 from https://www.generali.si/vizija-poslanstvo-vrednote

8. Generali zavarovalnica d. d. (2019). *Zaupaj v svojo pot - predstavitev zavarovalnice Generali*. Retrieved from https://www.generali.si/documents/180316/183247/Brošura+-+Zaupaj+v+svojo+pot/aa7dad72-d971-4ec5-9ba1-77e9929a3145

9. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

10. Hassan, M. ul. (n.d.). *VGG16 – Convolutional Network for Classification and Detection*. Retrieved August 2, 2020 from https://neurohive.io/en/popular-networks/vgg16/

11. Haykin, S. (2009). *Neural Networks and Learning Machines* (3rd ed.). New York: Prentice-Hall.

12. Imagenet. (n.d.). *About ImageNet.* Retrieved August 2, 2020 from http://www.image-net.org/about-overview

13. Ivanjko, Š., Ivanjko, S., Ivanjko, L., & Ihanec, K. (1999). *ABC zavarovalništva.* Maribor: Založba Kapital – Neto d. o. o.

14. Jakupović, E. (2018, 8 November). *Štirje stadiji na poti do umetne inteligence*. Retrieved from https://ikt.finance.si/8941038/Stirje-stadiji-na-poti-do-umetne-inteligence

15. Jupyter. (n.d.). *Jupyter – About Us*. Retrieved August 4, 2020 from https://jupyter.org/about

16. Keras. (n.d.). *About Keras*. Retrieved August 4, 2020 from https://keras.io/about/

17. Kriesel, D. (2007). *A Brief Introduction to Neural Networks.* Retrieved from http://www.dkriesel.com/en/science/neural_networks

18. Lexico. (n.d.). *Overfitting definition*. Retrieved August 4, 2020 from https://www.lexico.com/definition/overfitting

19. MathWorks. (n.d.). *InceptionResNetV2*. Retrieved August 4, 2020 from https://www.mathworks.com/help/deeplearning/ref/inceptionresnetv2.html

20. Mikrocop. (n.d.). *Arhiviraj.si*. (interno gradivo). Retrieved March 10, 2020 from http://arhiviraj.si

21. Ojsteršek, V. (2005). *Zavarovalništvo: skripta za predmet*. Visoka komercialna šola.

22. OpenGenus. (n.d.). *Understanding the VGG19 Architecture*. Retrieved August 4, 2020 from https://iq.opengenus.org/vgg19-architecture/

23. Özsert Yiğit, G., & Özyildirim, B. M. (2018). Comparison of convolutional neural network models for food image classification. *Journal of Information and Telecommunication, 2,* 347–357.

24. Piccinini, G. (2004). The First computational theory of mind and brain: a close look at McCulloch and Pitts's "logical calculus of ideas immanent in nervous activity". *Synthese, 141* (2), 175-215.

25. Python. (n.d.). *About Python*. Retrieved August 4, 2020 from https://www.python.org/about/

26. Rozman, A. (2017, 21 March). *Reševanje zavarovalnih primerov*. (interno gradivo). Generali zavarovalnica d. d.

27. Rozman, M., & Simončič, M. (2020). *Navodilo za proces dela s sklepalnimi dokumenti*. (interno gradivo). Generali zavarovalnica d. d.

28. Shukla, N. (2020). *Business Process Automation (BPA): Where it Works and Key Benefits [Practical Examples Included]*. Retrieved August 26, 2020 from https://www.appypie.com/business-process-automation

29. Simončič, M., Hrvatin, S., & Siter, A. (2017). *Določitev poslovnega procesa*. (interno gradivo). Generali zavarovalnica d. d.

30. Sodstvo Republike Slovenije. (2018). *Izvršba – izterjava dolga po sodni poti*. Retrieved from https://nasodiscu.si/izvrsba

31. Svet kapitala. (2018, 26 November). *Robotska avtomatizacija kot orodje za izboljšanje procesov*. Retrieved from https://svetkapitala.delo.si/b2b/robotska-avtomatizacija-kot-orodje-za-izboljsanje-procesov/

32. Šker, T. (2010). *Osnove zavarovalništva.* Ljubljana: Zavod IRC.

33. Thackray, R. (1981). The stress of boredom and monotony: A consideration of the evidence. *Psychosomatic medicine*, *43* (2), 165-176.

34. Wikipedia. (n.d.). *Convolutional neural network*. Retrieved August 4, 2020 from https://en.wikipedia.org/wiki/Convolutional_neural_network#cite_note-Valueva_Nagornov_Lyakhov_Valuev_2020_pp._232–243-1

35. Wikipedia. (n.d.). *Deep learning*. Retrieved August 4, 2020 from https://en.wikipedia.org/wiki/Deep_learning#Deep_neural_networks

36. Wikipedia. (n.d.). *ImageNet*. Retrieved August 4, 2020 from https://en.wikipedia.org/wiki/ImageNet

37. Wikipedia. (n.d.). *Project Jupyter*. Retrieved August 4, 2020 from https://en.wikipedia.org/wiki/Project_Jupyter#Jupyter_Notebook

38. Wikipedia. (n.d.). *Python (programming language)*. Retrieved August 4, 2020 from https://en.wikipedia.org/wiki/Python_(programming_language)#cite_note-AutoNT-7-28

39. Zavarovalnica Triglav. (2015, April). *Drobni tisk: Regresni zahtevek*. Retrieved from https://vsebovredu.triglav.si/dom/regresni-zahtevek

# APPENDICES

# APPENDIX 1: POVZETEK V SLOVENŠČINI (SUMMARY IN SLOVENE LANGUAGE)

Želja po samozaščiti in zaščiti svoje družine je bila vedno del človekovega življenja. Organizirana skupna zaščita je predstavljala večjo varnost, zlasti če vsi prispevajo k načinu zaščite pred tveganji, s katerimi se soočajo – zavarovanju. V današnjem svetu je zavarovalništvo gospodarska in družbena dejavnost, katere namen je spodbuditi ekonomsko varnost. Definiramo ga kot sredstvo za zaščito pred različnimi situacijami, ki povzročajo finančno izgubo.

Od začetkov organizacije zaščite – zavarovanja je tehnologija izjemno napredovala. Ljudje nenehno izdelujejo nove izume, ki poenostavljajo in nadomeščajo njihovo delo na vseh področjih, vključno z zavarovalništvom. Izkoriščanje funkcij, ki jih lahko opravljajo računalniki, je eden izmed načinov poenostavitve dela, ki se uporablja že nekaj desetletij. Ena od novosti tehnologije so umetne nevronske mreže, ki so bile izdelane po modelu bioloških nevronskih mrež. Slednje slovijo po svojih izjemnih sposobnostih ter po lastnosti, da se skozi celo življenje učijo od okolja.

Umetne nevronske mreže so zgrajene kot algoritmi za reševanje raznolikih težav. Tako kot biološke nevronske mreže se tudi umetne lahko učijo in izboljšujejo. Ta način učenja nevronskih mrež imenujemo globoko učenje. Globoko učenje se uporablja na številnih področjih, vključno z zavarovanjem. V mojem magistrskem delu sem se usmerila v uporabo nevronskih mrež in globokega učenja za pregledovanje prisotnosti podpisa stranke na zavarovalnih policah motornih vozil. Podpis stranke na pogodbah je izjemno pomemben, saj dokazuje, da se stranka strinja z vsem, kar pogodba vključuje. To pomeni, da razume, kateri elementi so všteti v polico in kateri niso in pod kakšnimi pogoji se zavarovanje upošteva, ter izkaže soglasje k višini in načinu plačila premije. V primeru konflikta med podjetjem in stranko, kjer je potreben sodni postopek, lahko nepodpisana zavarovalna polica prevesi rezultat v prid stranki. To lahko vodi do izgub dohodka podjetja, česar se le-to želi ubraniti.

Glavni cilj mojega magistrskega dela je bil izdelava modela globokega učenja, ki bo razpoznal prisotnost podpisa stranke na zavarovalnih policah motornih vozil. Poleg tega je bil moj cilj tudi raziskati izzive in priložnosti modela, ter koristi, ki bi jih imelo podjetje ob uporabi tovrstnega modela.

Model, ki sem ga naredila, je dosegel 94.79% natančnost pri razpoznavi podpisa stranke. S tem sem ugotovila, da je model možno narediti že z relativno majhno podatkovno množico – 1920 zavarovalnih polic ter da se s številom podatkov model izboljšuje. Uporabila sem več vrst prej naučenih modelov kot podlago za moj model, vendar sta se le dve izkazali za uspešni – VGG16 ter VGG19.

Ugotovila sem, da bi avtomatizacija pregledovanja prisotnosti podpisa stranke na zavarovalnih policah podjetju prinesla časovno in ekonomsko učinkovitost. Čeprav pregled vsakega podpisa traja le nekaj sekund, to ob veliki količini polic predstavlja omembe vredno količino časa. Poleg tega bi avtomatizacija preverjanja podpisa stranke pomagala med samim postopkom sklepanja zavarovalnih pogodb. Zavarovalniški asistenti, ki trenutno opravljajo to nalogo, bi se lahko osredotočili na zahtevnejše naloge, kot so pregledovanje pogojev na polici, višine premije, ugotavljanje, ali so kakšne napake na polici ipd. Tako bi duhamorne naloge, kot je preverjanje podpisa stranke, prepustili računalniku. To bi preprečilo monotonost na delu; poleg tega bi omogočilo še boljšo kakovost drugih opravljenih nalog.

Čeprav model prinaša nekaj izzivov, kot so na primer možnosti za kibernetske napade, napake v programski ali strojni opremi in morebitno preveliko odvisnost od tehnologije, ponuja veliko pomembnih priložnosti, ki jih lahko izkoristimo. Ponuja večjo hitrost in zanesljivost naloge preverjanja podpisa stranke, manj napak in tveganj pri tej nalogi, nižje obratovalne stroške in večjo delovno etiko med zaposlenimi. Izdelan model bi se lahko tudi nadgradilo tudi tako, da bi prepoznaval podpise tudi na ostalih zavarovalnih policah, samodejno reklamiral nepodpisano zavarovalno polico ali celo prepoznaval škodo na slikah zavarovanih lastninah. Zaposleni opravljajo tudi veliko drugih nalog, ki bi jih lahko avtomatizirali in s tem poenostavili poslovne procese. Tako področje nevronskih mrež in globokega učenja prinaša veliko možnosti za uporabo v zavarovalništvu, ki še niso izkoriščeni. Moj model globokega učenja, predstavljen v magistrski nalogi, podjetju prinaša cenovne in časovne prihranke ter prav tako večjo učinkovitost delovanja podjetja na tem področju dela.

# APPENDIX 2: INTERVIEW

**I. Izkušnje v zavarovalništvu in procesu Upravljanja z zavarovalnimi pogodbami**

1. Kdaj ste začeli karierno pot v zavarovalništvu?
2. Koliko časa že delate v kontroli zavarovalnih pogodb?

**II. Razvoj poslovnega procesa – vpliv avtomatizacije, uvedbe orodja za upravljanje poslovnih procesov (Ultimus)**

1. Kako je izgledalo (potekalo) delo procesa Upravljanje z zavarovalnimi pogodbami takrat, ko ste začeli na tem področju?
2. Digitalizacija, avtomatizacija so izzivi, s katerimi se danes srečujejo vsa podjetja. Kako se to pozna pri delu v vaši ekipi?
3. Kako se spreminjajo znanja, ki jih potrebujete v ekipi?
4. Kako te spremembe vplivajo na kadre (se potrebuje manj kadra; le-ti pa morajo imeti več znanja)?

**III. Podpis zavarovalca na zavarovalni pogodbi**

1. Kaj so največji problemi, s katerimi se podjetje sooča glede nepodpisanih polic?
2. Kakšne kriterije uporabljate pri kontroli podpisa (ali poskušate razbrati vsebino podpisa; ali samo pogledate, da je nekaj napisano; kako daleč od predvidenega mesta je lahko podpis)?
3. Koliko časa vzame kontrola podpisa vaši ekipi (vključno z izdajo reklamacije)?

**IV. (Avtomatizacija kontrole podpisa zavarovalca na polici) Uporaba modela nevronske mreže za kontrolo podpisa**

1. Kako ocenjujete doprinos avtomatizacije kontrole podpisa (časovna, ekonomska učinkovitost)?
2. Kakšne so bo vašem mnenju prednosti avtomatske kontrole?
3. Ali menite, da bo model zanesljivejši od človeka?
4. V kašno delo boste usmerili zaposlene za pridobljeni čas?
5. Ali vidite tudi kakšne slabosti (pomanjkljivosti) ali celo nevarnosti (tveganja) avtomatizacije?

**V. Nadaljnje priložnosti modela**

1. Kakšno je vaše mnenje glede uporabnosti modela? Kakšne priložnosti prinaša?
2. V katerih procesih zavarovalnice bi lahko še uporabili model za prepoznavo podpisa?
3. Ali je v vašem procesu še kakšna aktivnost, ki bi jo lahko prevzel model nevronske mreže?

**VI. Imate še kakšno opažanje, misel ali mnenje o modelu in njegovem delovanju v podjetju?**