Prediction of the Influenza Virus Propagation by using different Epidemiological and Machine Learning Models

Bachelor Thesis

in partial fulfillment of the requirements to obtain the academic degree
Bachelor of Science (B.Sc.) in the course of studies
Medical Informatics

submitted by

Carina Schmidt

February 20, 2019

1st Supervisor:	Prof. Heinrich Krayl, Heilbronn University
2nd Supervisor:	Ahmed Bounekkar, Polytech Lyon, Université Lyon 1

Abstract

In this bachelor thesis, different models for predicting the influenza virus are examined in more detail.

The focus is on epidemiological compartmental models, as well as on different Machine Learning approaches.

In particular, the basics chapter presents the SIR model and its various extensions. Furthermore, Deep Learning and Social Network approaches are investigated and the applied methods of a selected article are analysed in more detail.

The practical part of this work consists in the implementation of a Multiple Linear Regression model and an Artificial Neural Network. For the development of both models the programming language Python was chosen using the Deep Learning Framework Keras.

Tests were performed with real data from the Réseau Sentinelles, a French organisation for monitoring national health.

The results of the tests show that the Neural Network is able to make better predictions than the Multiple Linear Regression model.

The discussion shows ideas for improving influenza prediction including the establishment of a worldwide collaboration between the surveillance centres as well as the consolidation of historical data with real-time social media data.

Therefore, this work consists of a state-of-the art of models regarding the spread of influenza virus, the development and comparison of several models programmed in Python, evaluated on real data.

Zusammenfassung

In dieser Bachelorthesis werden verschiedene Modelle zur Vorhersage der
Ausbreitung des Influenzavirus genauer untersucht.

Der Schwerpunkt liegt dabei auf der Analyse verschiedener epidemiologischer
Kompartiment-Modellen und Machine Learning Ansätzen.

Insbesondere wird im Grundlagenkapitel das SIR-Modell und seine verschiedene
Erweiterungen vorgestellt. Des Weiteren werden Deep Learning und Social
Network Ansätze untersucht und die angewandten Methoden eines ausgewähl-
ten Artikel ausführlicher beleuchtet.

Der praktische Teil dieser Arbeit besteht aus der Implementierung eines Multi-
plen Linearen Regressionsmodells, sowie eines Künstlichen Neuronalen Netzes.

Für die Entwicklung der beiden Modelle wurde die Programmiersprache Python
unter Benutzung des Deep Learning Frameworks Keras verwendet.

Tests wurden mit reellen Daten des Réseau Sentinelles, einer französischen
Organisation zur Beobachtung des nationalen Gesundheitszustandes, durchge-
führt.

Die Ergebnisse der Tests zeigen, dass das Neuronale Netz in der Lage ist,
bessere Vorhersagen zu machen, als die Multiple Lineare Regression.

In der Diskussion werden Ideen zur Verbesserung der Influenzavorhersage
vorgeschlagen. Diese beinhalten die Kollaboration zwischen Überwachungszen-
tren weltweit, sowie die Zusammenführung historischer Überwachungsdaten
mit Echtzeitdaten sozialer Netzwerke.

Diese Arbeit besteht daher aus einem State-of-the-art von Modellen zur Aus-
breitung der Influenza, sowie der Entwicklung, als auch dem Vergleich mehrerer
in Python programmierter Modelle, die anhand von realen Daten ausgewertet
wurden.

Résumé

Dans ce mémoire de fin d'étude, différents modèles permettant de prédire la propagation du virus de la grippe sont examinés en détail.

L'accent est mis sur les modèles épidémiologiques compartimentaires, ainsi que sur différentes approches d'apprentissage machine.

En particulier, un chapitre est consacré à l'état de l'art du domaine, c'est à dire au modèle SIR et à ses différentes extensions.

En outre, les approches d'apprentissage profond et de réseaux sociaux seront étudiées et un modèle récent sera analysé plus en détail.

Enfin, la partie expérimentale de ce travail consiste en la mise en œuvre d'un modèle de régression linéaire multiple et d'un modèle d'apprentissage par réseau de neurones profond. Pour le développement des deux modèles, le langage de programmation Python a été utilisé, et les approches implémentées en utilisant le framework Keras, spécifique au développement de modèles d'apprentissages basés sur le gradient.

Les tests ont été réalisés sur des données réelles du Réseau Sentinelles, un organisme français de surveillance de la santé nationale. Les résultats des expérimentations montrent que le réseau neuronal est capable de meilleures prédictions que la régression linéaire multiple.

Dans la discussion, différents axes d'amélioration des modèles présentés pour la prédiction de la propagation de la grippe sont proposés. Les idées comprennent l'établissement d'une collaboration entre les centres de surveillance du monde ainsi que l'agrégation des données historiques avec les données des réseaux sociaux en temps réel.

Ce travail consiste donc en un état de l'art des modèles permettant de modéliser la propagation de virus, et à l'élaboration et comparaison de plusieurs modèles d'apprentissage programmés en Python, évalués sur des données réelles.

Acknowledgements

I would first like to thank my thesis advisor Prof. Heinrich Krayl of Heilbronn University for encouraging me to go to Lyon, the organisation and the support before and during the thesis.

I would also like to acknowledge Ahmed Bounekkar of Polytech at Université Claude Bernard Lyon 1 for making my internship with this interesting topic at laboratoire ERIC possible and for always being available for any question.

I am very grateful to the members of Laboratoire ERIC for welcoming me in the team and for their support.

Moreover I want to thank my great friends from all over the world that I had the pleasure of getting to know in Lyon. Thank you for being part of my life and making my stay in Lyon unforgettable!
Thank you, Gebrüder Jim from Churchtown, for always being there for me!
Jim T, grazie mille for accompanying through my whole studies, for our freakiness and for reading this thesis!
Furthermore I would like to thank my parents and my entire family for making my stay in Lyon possible, the patience, the support and for enduring me during my whole study time.
Last but not least, I want to thank Tom for firmly standing behind me, encouraging me and always believing in me.

Thank you very much! Vielen Dank! Merci beaucoup! Grazie mille! Muito obrigada! Tusen takk! Muchas gracias!

List of Abbreviations

AI	Artificial Intelligence
ANN	Artificial Neural Network
BP	Back-Propagation
CDC	Centers for Disease Control and Prevention
CNN	Convolutional Neural Network
DL	Deep Learning
ECDC	European Centre for Disease Prevention and Control
EISN	European Influenza Surveillance Network
GA	Genetic Algorithm
GD	Gradient Descent
GFT	Google Flu Trends
ILI	Influenza-like illness
LS	Least Squares
MAE	Mean Absolute Error
MSE	Mean Squared Error
MCMC	Markov chain Monte Carlo
ML	Machine Learning
MLP	Multi-Layer Perceptron
MRM	Multiple Regression Model
NN	Neural Network
ReLU	Rectified Linear Unit
RNN	Recurrent Neural Network
SI	Susceptible-Infected
SIR	Susceptible-Infected-Recovered
SIS	Susceptible-Infected-Susceptible
SVM	Support Vector Machine
VGP	Vanishing Gradient Problem
WHO	World Health Organization

Contents

1 Introduction

The focus of this thesis was centred on the development of a tool to predict the propagation of influenza. In particular, the task was to find out how the propagation can be best predicted by comparing different epidemiological, deterministic, stochastic and Machine Learning approaches. In the end two models were selected to be implemented and tested.

1.1 Epidemiology

As defined by the World Health Organization, "epidemiology is the study of the distribution and determinants of health-related states or events (including disease), and the application of this study to the control of diseases and other health problems. Various methods can be used to carry out epidemiological investigations: surveillance and descriptive studies can be used to study distribution; analytic studies are used to study determinants".[WHO18]

The word "epidemic" originates from the Greek: "epi", which means upon or above, and "demos", which means people. An epidemic is a highly concentrated occurrence of a disease, especially of an infectious disease, limited in time and geographically. It spreads quickly and effects a large number of individuals within a population at the same time.[Fle18]

To distinguish between epidemic and pandemic, a pandemic describes the spread of a particular disease, which is unlimited in location and affects a large proportion of the population over large areas and continents. This is the case with some influenza viruses and globally occurring diseases such as HIV or tuberculosis. In contrast, an "endemic" is the local spread of a disease that occurs only in a certain population or in a certain area, for example Cholera or Malaria. In common parlance, "pandemic" and "endemic" are often wrongly considered as "epidemic".[Fle18]

A basic concept in epidemiology is the existence of thresholds. These are critical values for quantities such as population size or vector density that must be exceeded in order for an epidemic to occur. The infectious contact number, which is the average number of contacts of an infective human during his infectious period, is identified as the threshold quantity which determines the behaviour of the infectious disease.[Fle18]

In general, a distinction is made between three different types of an epidemic: First, the spread by contagion, it is the direct transmission between humans, for example from a contaminated to a susceptible individual. The individual is therefore the only carrier and transmitter of the virus. Second, the spread generated by the use of air, rail and maritime transport which causes a hierarchical spread. The propagation takes place either from a small town to a large city (upward spread) or vice versa (downward spread). Third, there is also a combination of contagion and hierarchical spread.[Bou14]
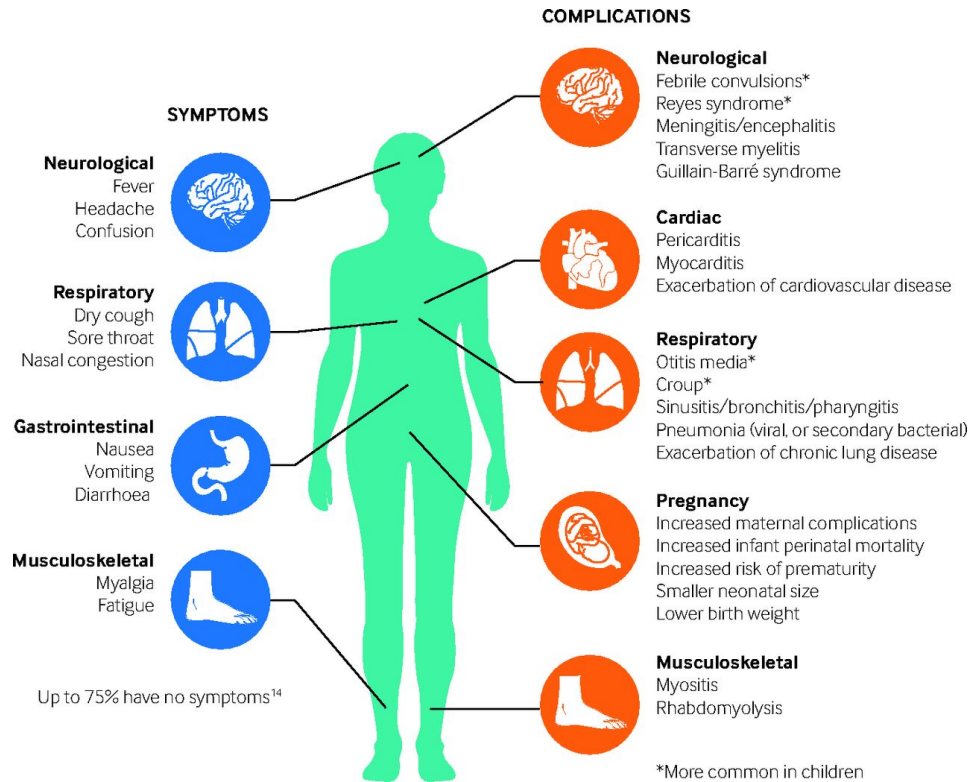
Figure 1.1: Symptoms and complications of Influenza

Epidemiological modelling has been used in the evaluation of prevention and control programmes. Understanding the phenomenon of virus transmission in the human population is a fundamental issue in epidemiology. Traditional epidemiological models assume that the transmission of an infectious in a homogeneous host population increases with the number of individuals and with the risk of a random infection. In epidemiology and infectious disease studies, we try to identify infectious agents and understand their way of propagation.[Bou14] The application of mathematical modeling to the spread of epidemics has a long history and was initiated by Daniel Bernoulli's work on the spread of smallpox in 1760[Huf04].

1.2 Influenza Virus

Influenza is a highly contagious respiratory disease that causes seasonal epidemics with significant morbidity and mortality. Influenza, commonly known as "the flu" is caused by different virus types. There are four types of influenza viruses: A, B, C and D, but only type A and B viruses cause clinically important human diseases and seasonal epidemics. Influenza A viruses cause the most severe disease and are the predominant cause of seasonal epidemics and pandemics in human populations [Ghe16].

Colloquially, the term "Grippe", which means influenza in the German-speaking part, is often misleadingly used for a common cold or an influenzial infection

("grippaler Infekt" or "Erkältung"). However, these two diseases differ in their symptoms and aetiopathology. An influenzial infection is characterised by the first symptoms, such as sneezing, runny nose and a mild feeling of illness. Sore throat, cough, headache and body aches as well as mild fever are possible symptoms in the further course of the disease, but after about a week, most of the complaints subside. In contrast, an influenza virus infection is characterized by a sudden onset of fever, cough, headache, myalgia, severe malaise, sore throat and nasal congestion. With a flu, one feels miserable in comparison to a cold and is ill for more than one week. Gastrointestinal symptoms including nausea, vomiting and diarrhoea are also common. In Figure 1.1, the symptoms as well as the possible complications are pointed out. The incubation period of influenza, this means the time from infection to development of symptoms, is one to four days.

1.2.1 Influenza surveillance

The European Influenza Surveillance Network (EISN) is coordinated by the European Centre for Disease Prevention and Control (ECDC). The network combines epidemiological and virological surveillance of influenza to provide decision makers and public health experts in EU member states with the information required to better assess influenza activity in Europe and take appropriate action. Finally, EISN aims to contribute to reducing the burden of disease associated with influenza in Europe. Weekly analysis of influenza surveillance is available at Flu News Europe, a joint output from ECDC and the World Health Organization (WHO) Regional Office for Europe.[EU19]

"GrippeNet.fr" is the French subproject participating at the European project "Influenzanet.eu" that includes 40000 volunteers from ten different countries. To find out more about influenza, it is necessary to have data. To obtain this data, the help of everyone is needed. "GrippeNet.fr" is an online surveillance system of Metropolitan France. In contrary to other surveillance systems, it monitors the data in real-time directly from the volunteers. Thanks to the participation of volunteers, the fast and flexible system allows the comparison of data with the other participating countries.[Sen19a]

"GrippeWeb" is a project of Robert-Koch-Institut and is the first web portal in Germany to monitor the activity of acute respiratory diseases using information from the population itself. With the help of this data, the information of the "Arbeitsgemeinschaft Influenza" (AGI), the Working Group on influenza, with which GrippeWeb cooperates closely, is supplemented. Since AGI monitors the activity of acute respiratory diseases at the level of primary medical care, the supplementation is possible. GrippeWeb, on the other hand, records what proportion of the total population is affected by an acute respiratory infection week after week and how many people have visited a doctor's office with such an illness.[Ins19]

1.2.2 Prevention of Influenza

Vaccination is the most effective means of preventing influenza and its complications. Immunity developed in one influenza season may not provide protection in future years mainly because of changes in circulating strains, waning immunity and antigenic drift, which are minor changes that occur in virus proteins between influenza seasons. Influenza vaccines are updated annually to include the viral strains that are predicted to circulate in winter.[Ghe16]

During a wave of influenza, the basic rules of hygiene have to be observed in general and especially in the event of illness in the immediate vicinity, such as in the family, due to the transmissibility of influenza viruses. In particular, the spread of the pathogens should be reduced by wearing a mouthguard, washing hands frequently with normal cleaning soaps, disinfecting contaminated surfaces and avoiding too close contact with the infected person. Influenza patients are often isolated as a prophylaxis against exposure during inpatient admission, usually in cohort isolation. [Fle18]

While vaccinations only strengthen adaptive immunity, vitamin D plays an important role in innate immune defence and adaptive immunity. Its active form intervenes in various ways in the immune system and promotes the formation of antimicrobial peptides. Various studies show that the susceptibility to influenza in persons with vitamin D deficiency could be significantly reduced if their vitamin D levels were increased by daily doses. A deficiency is, among others, favoured by dark skin colour, various digestive disorders, increased stay in rooms and vehicles, darker or rainy seasons. A control of the vitamin-D-level is recommended and at values below 30 ng/ml, vitamin D should be supplied. However, the role of vitamin D in preventing influenza is not undisputed.[Mar17]

1.3 Motivation

1.3.1 Need for influenza modelling

Worldwide, annual epidemics are estimated to occur in approximately 3 to 5 million annual cases of severe illness and approximately 250.000 to 500.000 deaths. As mentioned before, to prevent the disease, the most effective way is vaccination. WHO recommends annual vaccination especially for high risk groups.[WHO18]

More than 110 countries in the world have a surveillance system that collects influenza viruses from clinical spaciments for detailed analysis and epidemiological information.[WHO18]

Since it is difficult to make a laboratory diagnosis for each suspected influenza patient, monitoring of Influenza-like illness (ILI) often serves as a substitute for influenza surveillance [Xue18]. The most lethal influenza pandemic in history

was the Spanish flu (H1N1), spread in 1918, costing an estimated 50 million lives worldwide. Further large influenza epidemics were the Asian flu (H2N2) in 1957, the Hong Kong flu (H3N2) in 1968, the Russian flu (H1N1) in 1977 and the flu pandemic (H1N1) known as the swine flu in 2009. Due to the flu's rapid mutations, it is necessary to take a vaccine every year to avoid falling sick or minimising the risk. But this practice is very criticed by health care and also not very convenient.[CDC18]

Due to the fact that influenza is a serious, sometimes deadly, widespread disease that causes severe impact and substantial economic loss to societies around the globe, an early and efficient influenza warning can save manpower and expense. Precisely and effective monitoring as well as predicting influenza activity can help to reduce the spread of the disease and minimise its risks.[Xue18]

Where and when influenza outbreaks will occur, how great the impact of the influenza season will be and when it will peak, varies from year to year and from season to season, making the preparation for and reaction to influenza very challenging. Influenza prediction can change that by providing the opportunity to look into the future and better plan ahead, possibly reducing the impact of the disease.

Influenza forecasts can be used to prepare for and prevent illness. Furthermore, the ability to more effectively plan for public health responses to seasonal epidemics and future influenza pandemics is possible. They can inform messaging to health care providers regarding influenza vaccination and antiviral treatment for patients. Moreover, predictions can help to prepare for an influx hospitalisation, helping inform the distribution and placement of health care staff and treatment resources. Finally, influenza forecasts can be used to guide community mitigation strategies, such as school closures.[CDC18]

1.4 Goals

The goals of this thesis are delineated in the following enumeration:

1. Demonstration of different models and approaches for modelling influenza

2. Choice and analysis of a model

3. Implementation of a model to predict influenza

4. Comparison of the results

1.5 Procedure and structure of the work

First, numerous articles regarding Influenza virus, epidemiological models and methods to predict the outbreak of a disease, were read.
Second, a state-of-the-art about different approaches for modelling and predicting influenza was written.

Third, different Machine Learning approaches, the programming language Python and the library Keras, were made familiar with.

Fourth, a Multiple Regression model in Python, as well as an Artificial Neural Network were implemented.

Furthermore, data for testing the model were chosen, thereafter the obtained results were interpreted and compared.

In the end, suggestions for better predict the outbreak of Influenza are proposed.

Hereinafter, this work is structured in the following chapters:

Chapter 2 comprises the most important Basics and the State of the Art to understand the project.

Chapter 3 describes the methods and technologies that where used to construct and implement the chosen models.

Chapter 4 shows the achieved results of the Multiple Regression Model and the Neural Network.

Chapter 5 discusses the results and provides an outlook for future research.

2 Basics and State of the art

2.1 Deterministic and Stochastic models

The propagation of an epidemic is defined as the growth in the number of infected individuals over time and space. It should be noticed that an epidemic tends to spread more quickly when each individual meets a certain number of other individuals in public places. Various models have been created to model epidemic spreading. Models are usually either stochastic or deterministic, although there are also mixed models.[Che05]

Deterministic models are based on known or hypothetical physical, mathematical laws, or laws from another discipline and contain only non-probabilistic components. The same input values will therefore always lead to the same results. Deterministic models are based on a system of ordinary or partial differential equations.[Mar15]

In contrast, stochastic models accept that there is a probability distribution associated with the input values, so that different output values can be obtained with the same data input. Stochastic models have one or more random (input) components and the output is also of random nature. They are either based on differential stochastic models or on Markovian processes. Multiple model simulations are required to analyse the output. Since stochastic models use probability density functions, they must be based on statistical theory.[Che05]

Furthermore, one can differentiate between static and dynamic. Static means that a system is described at one point in time and dynamic means that the system description changes in time. Moreover, one can distinguish continuous-time, where the model allows the system state to change at any time, and discrete-time, where changes are allowed only at distinct times.[Mar15]

In general, stochastic epidemic models are to be preferred over deterministic ones if their analysis is possible. On the one hand, a natural way to describe the spread of a disease is stochastic; one defines the probability of disease transmission between two individuals rather than stating certainly whether or not the transmission will occur. On the other hand, deterministic epidemic models rely on the law of large numbers. Furthermore, when considering extinction this can only be analysed with stochastic models [Gon15].

2.2 Compartmental models

The easiest way to model epidemic spread in populations is to classify people into different population groups or compartments. Compartmental models are governed by a system of differential equations that track the population as a function of time, stratifying it into different groups based on risk or infection

status. The models track the number of people in some or all of the following categories: Susceptible (S), Exposed (E), Infectious (I) and Recovered (R) [Dis18]. These abbreviations are used in the following sections. In epidemiology, a susceptible individual is a healthy person who is capable of contracting the disease. Exposed individuals have been infected with a pathogen, but due to the pathogen's incubation period, they are not yet infectious. An individual is infectious when he or she can spread the pathogen around him or her. Infected individuals have the disease and can spread it to every susceptible individual. Recovered people are either no longer infectious or "removed" from the population.[Mar15]

2.2.1 SI model

The simplest model to describe an epidemic propagation is the so called SI model. The model assumes that the total number of people is N. It divides the population in only two compartments. An individual is considered to go through two phases during the epidemic period. The impact of birth and death rates on the total number of people is not regarded [Li14]. An individual is first susceptible, then infectious. In case of an SI model, researchers made a strong hypothesis: as soon as a susceptible individual comes into contact with an infectious individual, he or she subsequently becomes infectious [Bou14]. The infectious rate β controls the rate of spread which represents the probability of transmitting the disease between a susceptible and an infectious individual. Recovery rate $\gamma = 1/D$ is determined by the average duration D of infection. The SI model without vital dynamics can be described with the following differential equations [Dis18]:

$$\frac{dS}{dt} = -\frac{\beta SI}{N} \tag{2.1}$$

$$\frac{dI}{dt} = -\frac{\beta SI}{N} = \beta I(1 - \frac{I}{N}) \tag{2.2}$$

2.2.2 SIS model

The SIS model additionally assumes that after an individual was infectious, he or she moves back to the susceptible class upon recovery. Healthy people contract an infection with the rate $a > 0$, infectious recover with the rate $b > 0$ by becoming susceptible again. Given that individuals remain susceptible after infection, the disease attains a steady state in a population, even without vital dynamics. The SIS model without vital dynamics can be described with the following equations [Dis18]:

$$\frac{dS}{dt} = -\frac{\beta SI}{N} + \gamma I \tag{2.3}$$

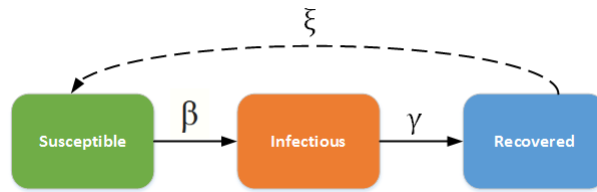$$\frac{dI}{dt} = \frac{\beta SI}{N} - \gamma I \tag{2.4}$$

Figure 2.1: SIR-SIRS model

Adding vital dynamics to the model, birth and death rate (μ and ν) are considered which leads to the following differential equations [Dis18]:

$$\frac{dS}{dt} = \mu N - \frac{\beta SI}{N} + \gamma I - \nu S \tag{2.5}$$

$$\frac{dI}{dt} = \frac{\beta SI}{N} - \gamma I - \nu I \tag{2.6}$$

W.H. Hethcote and P. Van Den Driessche propose a SIS model including births and natural mortality, that are related to the disease and exponential demographic structure. In this model they take a time lag into account for the infectious period and consider that the size of the population is variable. The authors hypothesize that the infectious period is constant for all individuals. The choice to vary the size of the population is justified by the fact that epidemiological and demographic processes interact and lead to new behaviours that do not occur when the population size is constant [Dis18].

2.2.3 SIR model

The epidemic modelling has its origin in 1927 and 1932 when W.O. Kermack's and A.G. McKendrick's described the well-known deterministic models in their papers [Ker27] [Ker32]. Their research and findings have had a major influence on the development of mathematical models for disease spread and are still common and relevant today, not only in the field of epidemics. Their primary idea was to divide the population into classes regarding their health situation. They defined three distinct compartments: S, I and R. In their SIR model, the susceptibles are healthy people who have never been infected and who can catch the disease. The time while people are infected and belong to the infected compartment is called infectious period. After an individual will be recovered from the disease, he or she will remain in the recovered compartment because they are assumed to be immune for their life time. In a closed population only a single influenza epidemic can occur because there is no inflow of susceptible individuals. Hence Figure 2.1 can be used to describe the order of the states individuals go through [Fre18], [Sta18].
The SIR model without vital dynamics can be described with the following ordinary differential equation (ODE) where $N = S + I + R$ is the total popu-

lation:

$$\frac{dS}{dt} = -\frac{\beta SI}{N} \tag{2.7}$$

$$\frac{dI}{dt} = \frac{\beta SI}{N} - \gamma I \tag{2.8}$$

$$\frac{dR}{dt} = \gamma I \tag{2.9}$$

The SIR model has been used as a basic principle for all consecutive epidemic models, particularly for influenza models. The simplest extension to the SIR model includes demographics, more precise the in- and outflow of individuals into the population [Cob09].

Regarding the birth (μ) and death rate (ν) and assuming that $\mu=\nu$ the SIR model with vital dynamics can be described with the following ODEs. In steady state $\frac{dI}{dt} = 0$.

$$\frac{dS}{dt} = \mu N - \frac{\beta SI}{N} - \nu S \tag{2.10}$$

$$\frac{dI}{dt} = \frac{\beta SI}{N} - \gamma I - \nu I \tag{2.11}$$

$$\frac{dR}{dt} = \gamma I - \nu R \tag{2.12}$$

$$\tag{2.13}$$

2.2.4 SIRS, SEIR and SEIRS models

In the **SIRS model**, the individual is susceptible, then infectious, then it has a temporary immunization and finally becomes susceptible again. In this case the immunisation is valid for a short period during the epidemic period.

In the **SEIR model**, the individual is susceptible and then infected but not infectious. Thereafter, it becomes infectious and finally has a permanent immunization (R).

In the **SEIRS model**, the individual is susceptible and then infected but not infectious. Thereafter, it becomes susceptible again. The SEIRS is a SEIR model with loss of immunity.[Bou14]

2.2.5 Extensions of Compartmental models

In addition, the arrival of globalisation has led some researchers to attempt a more refined approach in their models. When modelling an epidemic, mobility is a significant factor to consider. Mobility is the ability of an individual to move from one geographical area to another leaving the used means of transport out of consideration. Therefore, there is a significant risk for travellers

to contract new pathogens during transport or stay and to spread them upon their arrival.

In the late 1960s L. A. Rvachev developed the first spatial-temporal model of influenza by connecting a series of SIR models in order to construct a network model of linked epidemics [Rva68]. As described in [Cob09], he modelled the geographic spread of influenza in the former Soviet Union by using travel data to estimate the degree of cross-linking between epidemics in big cities. He and his colleagues Baroyan and Longini extended the network model and evaluated the effect of air travel on influenza pandemics in 1985 [Rva85] which laid the foundation for other modeling studies attaching importance to air travel on geographic spread[Cob09] .

Hufnagel et al.[Huf04] proposed a spatio-temporal model on two scales that describes the worldwide spread of infectious diseases and demonstrates that a forecast of the geographical spread of epidemics is possible. The standard SIR model, which is deterministic, does not account for fluctuations. These fluctuations are particularly important at the beginning of an epidemic when the number of infected is very small. Their model combines a stochastic local infection dynamics among individuals with stochastic transport in a worldwide network, taking into account national and international civil aviation traffic. They focused on mechanisms of the worldwide spread of infectious dynamics and the global travelling dynamics of individuals. Both constitutes of the model are treated on a stochastic level, taking full account of fluctuations of disease transmissions, latency and recovery on the one hand and geographical dispersal of individuals on the other hand.

El Kiha et al.[El 17] devised a multi-regions SEIRS discrete-time model to describe epidemics dynamics when an infection spreads from one region to another via air travels. They chose a grid of cells where each cell represents a region for giving an example of simulations. In the end, they showed the effectiveness of the travel-blocking vicinity optimal control approach when it is applied to one cell and when movements of infected people from the vicinity are restricted, so that they could keep the region with insignificant infection.

J. Saramäki and K. Kaski[Sar05] presented in their paper a model for the spread of randomly contagious diseases such as influenza. It is based on a stochastic SIR model on dynamic small-world networks, where randomly occurring long-range (LR) links are introduced in order to take into account the inherent randomness of spreading. They have derived equations for the epidemic threshold and spreading dynamics and shown that those match results of discrete time-step simulations.

A compartmental SIR model including two biomedical interventions has been developed by Coburn et al.. Susceptible individuals (S) can be vaccinated and infected individuals (I) can be treated with antiviral drugs (T). After having received the vaccination, the previous susceptible individuals are considered immune, whereas people having taken medication are considered less infectious in class "T".[Cob09]

Another extension regards two behavioural interventions: quarantine and isolation. Susceptible individuals can be quarantined and after recovery return to the class of susceptibles again. Infected, asymptomatic and not yet infectious individuals (E) can be quarantined. If they then develop symptoms and become infectious they can be isolated like the infected individuals [Cob09]. Different behavioural interventions have been analysed such as closing schools, quarantining individuals or travel restrictions. Ferguson et al.[Fer06] have determined in their paper that household quarantine could be more effective than closing schools.

2.3 Prediction of Influenza with Machine Learning approaches

Forecasting events is not new to medicine. In combination with modern Machine Learning, clinical data sources enable us to rapidly generate prediction models for thousands of similar clinical questions. From early-warning systems to imaging diagnostics, the potential applicability of these approaches is extensive. Many methods for predictions and classifications exist. The ability to predict influenza peak several weeks in advance allows timely preventive public health planning and interventions to be used to mitigate these outbreaks [Che17].

In the following subsections, fundamentals of Machine Learning are explained and different articles are presented.

2.3.1 Machine Learning vs. Deep Learning

Machine Learning (ML) is a subfield of Artificial Intelligence which deals with algorithms that allow computers to learn. The concept of these procedures is to learn the theory automatically from the data through processes of inference, learning from examples or model fitting. A model is defined with several parameters and the learning process is the implementation of a programme using training data to optimise the performance of the model [Att10]. Two of the most widely used ML methods are supervised and unsupervised learning.

Supervised learning algorithms deal with labeled data, so the algorithm receives the instruction of how the output based on a specific input should look like.

Unsupervised learning algorithms experience a dataset without any labels and are supposed to learn the whole probability distribution.[Att10]

ML algorithms are capable to solve everyday or very specific problems. In practice, however, problems often arise when there is either too little data or when there are too many dimensions of the data. Entropy-driven learning algorithms, such as decision trees, often become too complex in many dimensions, and vector-space-based algorithms such as the k-nearest neighbor algorithm are limited in their performance by dimensionality. [Aun18]
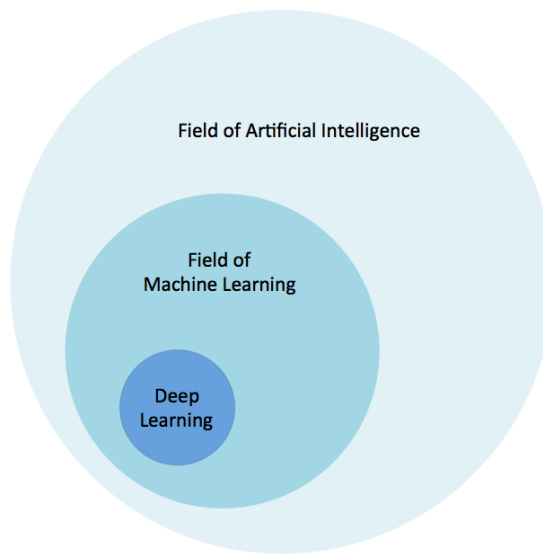
Figure 2.2: Classification of Deep Learning

2.3.1.1 Support Vector Machine

In ML, Support Vector Machine (SVM)s are supervised learning models with associated learning algorithms that analyse data used for classification and regression analysis. A SVM is a discriminative classifier formally defined by a separating hyperplane. In other words, given labeled training data (supervised learning), the algorithm outputs an optimal hyperplane which categorises new examples. A SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. In addition to performing linear classification, SVMs can efficiently perform a non-linear classification, implicitly mapping their inputs into high-dimensional feature spaces.
Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier.[Lia18]

2.3.1.2 Deep Learning

Deep Learning (DL) is a discipline of Machine Learning (ML) using Artificial Neural Networks. Figure 2.2 illustrates the relation between DL, ML and Artificial Intelligence (AI) [Pat17]. While the ideas for decision trees or k-means were developed from a certain mathematical logic, there is a model from nature for Artificial Neural Networks: A Biological Neural Network. DL is an approach for generating hierarchical representations from input data itself and mapping these to an output. It is a set of learning methods attempting to model data with complex architectures combining different non-linear transformations. The advantage of ANNs is the very deep abstraction of correlations between input data, the abstracted neuron values and the output
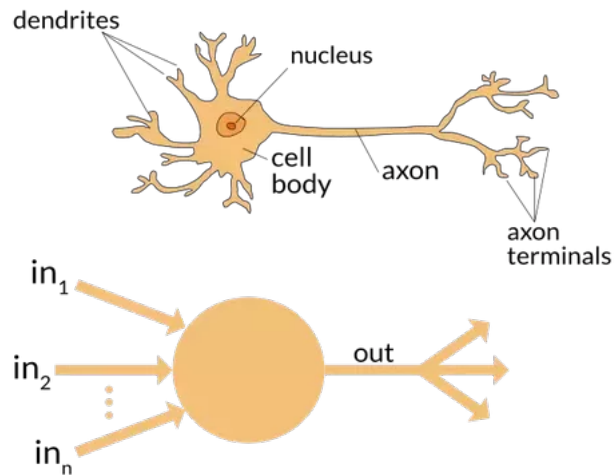
Figure 2.3: Biological and artificial neuron

data. This happens over several layers of the networks, which can solve very specific problems. From these facts the superordinate name is derived: Deep Learning. Deep Learning is used when other Machine Learning methods reach their limits. It has become one of the central development drivers in AI in the recent years for two reasons. First, because DL achieves particularly good results when large amounts of data (Big Data) are available to train the network. Secondly, because DL algorithms make it possible to represent intellectual and mental processes that had long been thought to be reserved only for humans [Tie18]. Almost all DL algorithms can be described as particular instances of a quite simple recipe: combine a specification of a data set, a cost function, an optimisation procedure and a model. The elementary components of DL are Neural Networks that are connected to form the Deep Neural Networks [Goo16].

2.3.2 Artificial Neural Networks (ANNs)

An ANN is a computational nonlinear model inspired by the the neural structure of the brain that is able to learn to perform tasks like classification, prediction, decision-making, visualisation, and others just by considering examples. It contains highly interconnected entities, called units or nodes [Goo16]. RNNs are Deep Learning technologies. They generally focuse on solving complex processes and are used for a variety of tasks including speech recognition, machine translation, computer vision, playing board games, social network filtering but also medical diagnosis. A typical neural network is a group of algorithms, which models data using neurons for ML[EDU18].

2.3.2.1 From biological to artificial neuron

The biological neuron is a nerve cell that provides the fundamental function unit for the nervous system of all animals. Neurons exist to communicate with
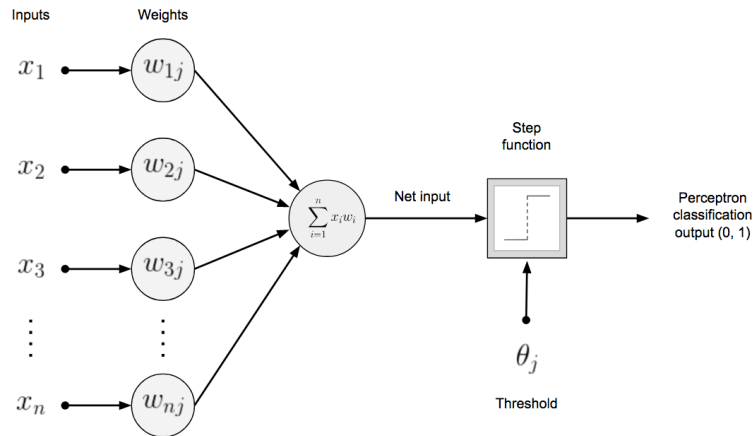
Figure 2.4: Structure of a perceptron

one another and pass electro-chemical impulses across synapses, from one cell to the next, as long as the impulse is strong enough to activate the release of chemicals across a synaptic cleft. The strength of the impulse must surpass a minimum threshold or chemicals will not be released. A neuron is made up of a nerve cell consisting of a soma, the cell body, that has many dendrites but only one axon. This single axon can branch hundreds of times. Axons are nerve fibers with a special cellular extensions that come from the cell body. Dendrites are thin structures that arise from the main cell body.

Figure 2.3 shows both the biological and the artificial neuron [Quo18]. One kind of artificial neurons is called perceptron, which was developed in the 1950s by the scientist Frank Rosenblatt. The perceptron is a linear model with a simple input-output relationship used for binary classification as illustrated in Figure 2.4. The figure shows that the number n of inputs is summed times their associated weights and then sent to a step function with a defined threshold. Characteristic of perceptrons is the use of the step function as an activation function which will output a real-valued single binary value (0 or 1), depending on the input. A perceptron can have several inputs $(x_1, ..., x_n)$ but produces only a single binary output. Weights $(w_1, ..., w_n)$ are real numbers, expressing the importance of the respective inputs to the output. The neuron's output (0 or 1) is determined by whether the weighted sum $\sum w_{i_j} x_j$ is less than or greater than the threshold value. The threshold is a real number, which is a parameter of the neuron.

By varying the weights, different models of decision-making can be received. The first layer of perceptrons makes very simple decisions by weighting the input evidence, while the second layer can make a decision at a more complex and more abstract level than perceptrons in the first layer. Other activation functions will be explained in the following sections. The perceptron training algorithm is considered a supervised learning algorithm. The precursor to the perceptron was the Threshold Logic Unit (TLU) developed by McCulloch and Pitts in 1943, which was able to learn the AND and OR logic functions. Both the TLU and the perceptron were inspired by the biological neuron [Pat17].
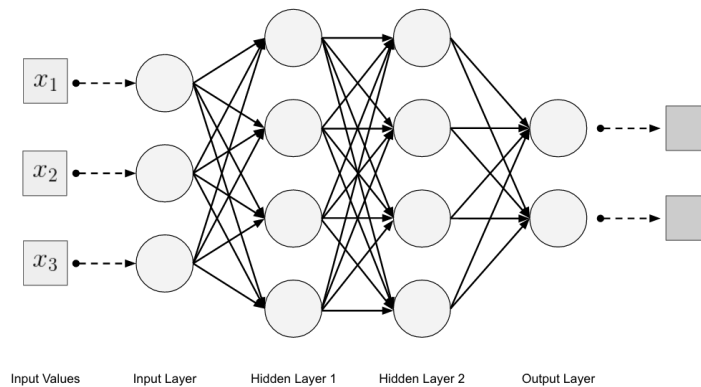
Figure 2.5: A Multilayer Neural Network topology

2.3.2.2 Components of ANNs

Like our brain is made up of millions of neurons, a Neural Network (NN) is just a composition of perceptrons, connected in different ways and operating on different activation functions. ANNs are organised in three interconnected layers: input, hidden and output layer. The input nodes provide information from the outside world to the hidden layer and are together referred to as the input layer. No computation is performed in any of the input nodes. The hidden nodes have no direct connection with the outside world, hence the name "hidden". They perform computations and transfer information from the input nodes to the output nodes. While a network will only have a single input layer and a single output layer, it can have multiple hidden layers. The output nodes are collectively referred to as the output layer and are responsible for computations and transferring information from the network to the outside world [Edu18].

Every neuron has weighted inputs (synapses), an activation function and one output. Synapses are the adjustable parameters that convert a neural network to a parameterised system [Wal19]. In Figure 2.5 the topology of a Multi-Layer Neural Network can be seen. The Feedforward Neural Network was the first and simplest type of an ANN. It contains multiple neurons arranged in layers. In a Feedforward network, the information moves in only one direction – forward – from the input nodes, through the hidden nodes and to the output nodes. There are no cycles or loops in the network. This property is different from Recurrent Neural Networks (RNNs), in which the connections between the nodes form a cycle. Examples of Feedforward networks are Single Layer Perceptron and Multi-Layer Perceptron (MLP), also called Deep Feedforward Neural Networks.

2.3.2.3 Activation functions

In an ANN, the activation function of a neuron defines the output of a neuron given a set of inputs. An activation function is biologically inspired by activ-

ity in our brains, where different neurons fire, or are activated, by different stimuli. Neurons get activated if the network input exceeds their threshold value.[Kri17]

An artificial neuron is a function f_j of the input $x = (x_1, ..., x_d)$ weighted by a vector of connection weights $w_j = (w_{j,1}, ..., w_{j,d})$, completed by a neuron bias b_j, and associated to an activation function ϕ, namely $f_j(x) = \sum(w_j * x) + b_j$. So, an artificial neuron calculates a "weighted sum" of its input, adds a bias and decides whether it should "fire" or not. Activation functions are used with Neural Networks to determine the output like yes or no. They map the resulting values e.g. in between 0 or 1 depending on the used function. The function can basically be divided into two types: linear or non-linear activation functions. The most commonly used activation functions are the Sigmoid or Logistic funtion, the Hyperbolic Tangent function (tanh) and the Rectified Linear Unit (ReLU) activation function. The important and used activation functions in this work are explained in Chapter 3.[Sch15]

2.3.2.4 Learning process of Neural Networks

There exist different types of learning but special focus in this thesis is placed on the supervised learning algorithm. When a NN is trained, it receives input data that will traverse the network. For each input data a corresponding output data should exist. Then, the result of the network is compared with the expected result. If there is a deviation in the result, all weights and biases in the perceptrons must be adjusted. A schematic representation of the process of a perceptron is shown in Figure 2.4. The output of the perceptron is negated and summed with the desired result. The resulting error is then squared to exclude mutual compensation of positive and negative errors. This results in the error E_R. The smaller the error E_R, the closer the result of the neurons at the target set by the teacher. To find the smallest error E_R, various approaches can be pursued.[Sch15]

The best known of all methods for solving unrestricted minimisation problems is the Gradient Descent (GD). Here the principle is pursued to find a relatively small error in the error mountains in a given number of steps. GD requires a cost function that wants to be minimised. Cost functions are used to monitor the error in predictions of a ML model. So, minimising the cost function basically means getting to the lowest error value possible or increasing the accuracy of the model. One increases the accuracy by iterating over a training data set while tweaking the weights and biases of the model. GD means that the direction is determined by the steepest gradient within the solution space mountain range. The problem becomes clear imagining the descent of a mountain where one always take the steepest path to get down as quickly as possible. In figure 2.6, there exists a local minimum where the gradient is zero. However, it is not the lowest loss one can achieve, which is the point corresponding to the global minimum. Now, if the weights are initialised at point A, then it will converge to the local minimum. The local minimum is
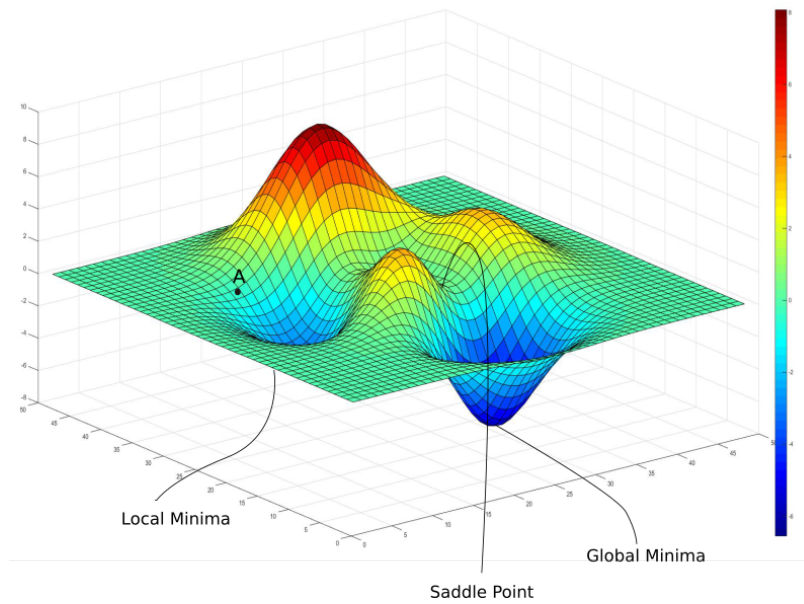
Figure 2.6: Cost function with local and global minima
[Pap18]

problematic because it is very difficult to find a way out of it with the classical
Gradient Descent. Neural Networks work according to the principle of error
minimisation through gradient descent. This always results in the danger of
running into a local minimum in the error mountains.[Rud16]

2.3.2.5 Recurrent Neural Networks

Recurrence is the process of a neuron influencing itself by any means or by any
connection. RNNs do not necessarily have explicitly defined input or output
neurons [Kri17]. In order to process information with a sequential context such
as text series or time series, RNNs have become the method of choice [Pal17].
RNNs are well suited for dealing with sequences. As diffusion episodes could
be seen as sequences, RNNs could be used to consider the history of events
to predict the future events. However, a direct application of RNNs does not
perform well for modelling diffusion because the process is better considered
as a tree.

The most simple RNN were developed in the 1980's, a hidden layer at time t
depends on the entry at time t, x_t but also on the same hidden layer at time
$t - 1$ or on the output at time $t - 1$.
We therefore have a loop from a hidden layer to itself or from the output to
the hidden layer.
RNN at the first glance, may seem very different from a classical NN. In
fact, this is not the case, as they can be considered multiple copies of the
same network, each passing information to its successor. The success of DL in
different fields also inspires some end-to-end representation learning framework

for cascade prediction.

Li et al.[Li17] proposed DeepCas, an end-to-end predictor of information cascades which transforms the cascade graph as node sequences through random walk and learns the representation of each cascade under a deep learning framework. It can automatically learn the valuable representations from data and is emerging with the success of representation learning in other areas. DeepCas ignores the temporal information for popularity prediction.

2.4 Methods of H. Xue et al.

In this section the work of H. Xue et al.[Xue18] and a selected part of the methods they proposed in their article "Influenza activity surveillance based on Multiple Regression Model and Artificial Neural Network" , is presented. First, the basic concepts of Least Squares, Multiple Regression, Back-Propagation and the Genetic Algorithm are explained, to make the models comprehensible.

2.4.1 Least Squares Method

The Least Squares (LS) method is a form of mathematical regression analysis that finds the line of best fit for a data set, providing a visual demonstration of the relationship between the data points. The basic problem is to find the best fit straight line $y = ax + b$ given that, for $n \in \{1, ...N\}$, the pairs (x_n, y_n) are observed [Mil06].

2.4.2 Multiple Regression Model

Regression just points the fact out that a model is continous and that for any inputs, you can get an output. The Multiple Regression Model (MRM) is an extension of the simple Linear Regression. While the simple Linear Regression regards only one dependent and one independent variable, the MRM adds more independent variables to the model. Multicollinearity and Overfitting are the two problems that may arise regarding a MRM. Adding more independent variables to a Multiple Regression procedure does not mean the regression will necessarily be better or offers better predictions. In fact, it can make things worse, which is called overfitting. According to [Uni19b] overfitting is "the production of an analysis that corresponds too closely or exactly to a particular set of data, and may therefore fail to fit additional data or predict future observations reliably". An overfitted model is a statistical model that contains more parameters than can be justified by the data [Uni19a]. The addition of more independent variables creates more relationships among them. So not only are the independent variables potentially related to the dependent variable, they are also potentially related to each other. The described relation between the variables is called multicollinearity. The ideal is for all independent

variables to be correlated with the dependent variable, but not with each other.

2.4.2.1 Back-Propagation

Back-Propagation (BP) is an algorithm for training a Neural Network. This algorithm was originally introduced in the 1970s. But only since the paper by David Rumelhart, Geoffrey Hinton and Ronald Williams [Rum85] was published in 1985, it became famous. They described several Neural Networks where Back-Propagation works faster than earlier approaches to learning, making it possible to use NNs to solve problems which had previously been unsolvable [Nie15]. BP is a supervised learning algorithm for training Multi-Layer Perceptrons. It is the recursive application of the chain rule along a computational graph to compute the gradients of all inputs [Kar16]. There are variants of training algorithms that use BP. The core of the concept is based on the adjustment of weights at each layer to reduce the resulting error while propagating the error. The BP algorithm looks for the minimum value of the error function in weight space using a technique called the delta rule or Gradient Descent. The weights that minimise the error function is then considered to be a solution to the learning problem [Edu18]. A Back-Propagation Neural Network (BPNN) is a Multi-Layer Feed Forward Network trained by an error inverse-propagation algorithm. The basic idea is to study the sample pair (input and expected output) and send the input of the sample to the BPNN to calculate the actual output. If the error between the actual output and the expected output does not meet the accuracy requirement, the error is propagated backwards to adjust the weights w and the threshold b so that the error between the output of the network and the expected output is gradually reduced until the accuracy requirement is satisfied [Ert18]. The BPNN is one of the most widely applied ANNs. On the one hand, its structure is rigorous, it has strong operability, and it can realise highly non-linear mapping. On the other hand, it is disadvantaged by slow learning convergence, elusive network topology, and a tendency to fall into the local minimum [Xue18].

2.4.3 Genetic Algorithm

J. Holland introduced the Genetic Algorithm (GA) in 1962 based on the concept of Darwin's theory of evolution. Afterwards his student extended it in the 1989 and genetic algorithms were applied to a wide range of subjects [Gol88]. GAs are commonly used to generate high-quality solutions to optimisation and search problems by relying on bio-inspired operators such as mutation, crossover and selection. GAs are a particular class of evolutionary algorithms that use techniques inspired by evolutionary biology such as inheritance, mutation, selection, and crossover. The evolution usually starts from a population of randomly generated individuals and happens in generations. In each generation, the fitness of every individual in the population is evaluated, multiple individuals are selected from the current population, based on their fitness,

$$ILI_{i,t} = \sum_{k=1}^{P} \chi_k X_{i,t-k} + \tau_t. \qquad (1)$$

$$ILI_{i,t} = \beta_1 X_{i,t} + \sum_{j \neq i, j=1}^{N} \lambda_j \omega_{i,j} X_{j,t} + \nu_t. \qquad (2)$$

Figure 2.7: GFT regression model (model 1) and with weights (model 2)

$$ILI_{i,t} = \sum_{k=1}^{P} \mu_k ILI_{i,t-k} + \sum_{m=1}^{P} \delta_m X_{i,t-m} + \sigma_t. \qquad (5)$$

Figure 2.8: GFT + CDC regression model (model 3)

and modified to form a new population. The new population is used in the next iteration of the algorithm. The algorithm terminates when either a maximum number of generations has been produced, or a satisfactory fitness level has been reached for the population [Mit98]. The commonly used vocabulary is: "individual" which means any possible solution, "population" that covers a group of all individuals, "fitness" stands for a target function that is used to optimise, "trait" is the possible aspect or features of an individual and "genome" represents a collection of all chromosomes for an individual.

2.4.4 Synopsis of an article

In their article "Influenza activity surveillance based on Multiple Regression Model and Artificial Neural Network", H. Xue et al.[Xue18] describe different methods to predict the propagation of influenza using a GA-BP Neural Network algorithm. They applied a Back-Propagation Neural Network based on the GA to fit the parameters of the flu-prediction regression model and compared the fitting effect with those of LS. All of their proposed models were utilised to assess and predict influenza activity across ten regions of the USA. For optimisation, the seasonal characteristics of influenza incidence were investigated and prediction models for the high-flu and low-flu season were established. The results show that overall investigated regression models, Centers for Disease Control and Prevention (CDC) regression model (model 4) and weighted CDC regression model (model 5) are superior to other models. The CDC regression model uses historical ILI data to predict future trends which in the absence of a new mutation impacting the case, yields good prediction. Combining these two models, results in the most accurate possible predictions.

The first model, proposed by Xue et al. is the Google Flu Trends (GFT) Regression Model, where they used GFT data to construct the equation.

The second model includes weights to the first model (Figure 2.7). The third

$$ILI_{i,t} = \sum_{k=1}^{P} \alpha_k ILI_{i,t-k} + \varepsilon_t. \tag{3}$$

Figure 2.9: CDC regression model (model 4)

$$ILI_{i,t} = \sum_{k=1}^{P} \beta_k ILI_{i,t-k} + \sum_{j \neq i, j=1}^{N} \lambda_j \omega_{i,j} ILI_{j,t} + \theta_t. \tag{4}$$

Figure 2.10: CDC regression model with weights (model 5)

model (Figure 2.8) combines the GFT+CDC regression model that uses historical flu-outbreak information of the CDC and the GFT regression that is better to reflect disease outbreaks in real time.

The US Centers for Disease Control and Prevention routinely count weekly influenza-related cases among all outpatients in ten surveillance regions of the United States and release weekly influenza-transmission data from national and regional disease-surveillance systems. The authors preferred CDC data to GFT data to establish the flu-prediction model because GFT data use original CDC data and the re-use of GFT in a model would have led to a secondary error. Due to the fact that the spread and infection of flu are continuous in time, the data processed by the predictive model is also continuous in time. Because the flu is mostly spread by contact among individuals, the relationship between any two regions of flu-affected populations is affected by the relationship between the people inhabiting these regions. If there is an outbreak of influenza in a region with frequent population movements, regions in close association are likely to be impacted. They propose the CDC regression model (model 4) as follows: Figure 2.9.

For model 5 (Figure 2.10) they integrated influenza-level information from across various regions into the real-time prediction model to establish the weighted CDC regression model. Historical CDC data was used to obtain the correlation coefficient between each region as the weight of the relationship: the greater the correlation, the greater $w_{i,j}$ value.

Influenza is generally considered to have seasonal epidemic characteristics, with a higher rate of prevalence in winter and spring months compared to the rest of the year. Increased contact among individuals due to being confined indoors under inclement weather, as well as the body's weakened resistance to disease during environmental temperature fluctuations, are some of the reasons for variations in flu prevalence by season. The goal of LS regression analysis is to find a concrete form or functional expression and to establish and analyse the functional relationship between a response (dependent variable Y) and an important factor (independent variable X), that is, to obtain a suitable function of X with which to express Y. Furthermore, they showed in their results that the use of the GA-BP Neural Network is superior to the LS-method.
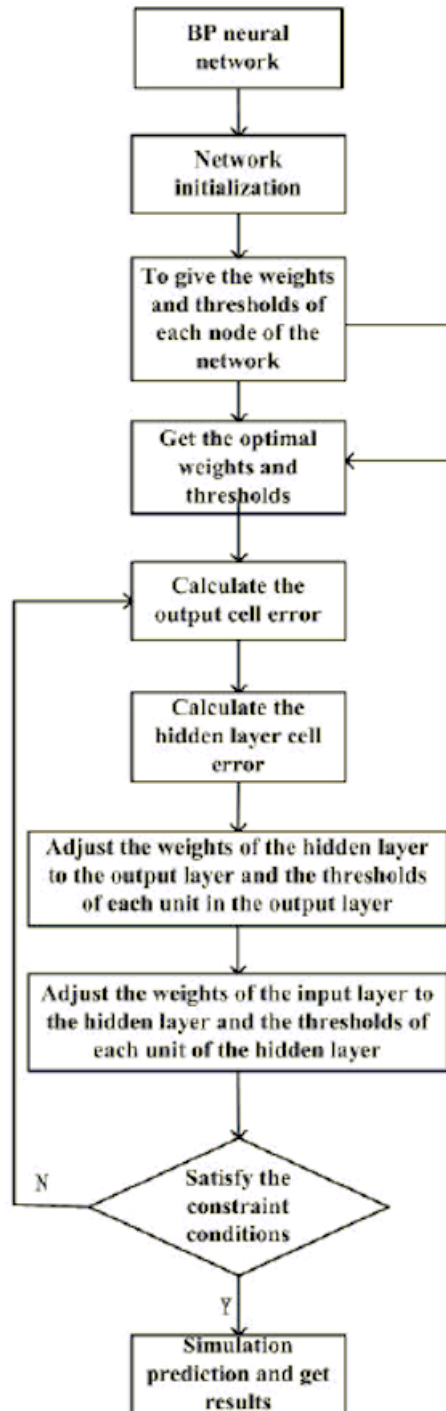
Figure 2.11: Topology of a Back-Propagation Neural Network

The models established in the authors' study provide a scientific reference for the future prediction and prevention of influenza. Reducing the risk of influenza via effective monitoring will benefit human health and prevent economic losses across the globe.

Below, I will focus on model 4 and model 5, as they have delivered the best prediction results among the five presented models. However, adding the seasonal fluctuations by considering the low-season and high-season of influenza led to even better results.

Model 4, the CDC regression model (Figure 2.9), predicts the number of persons suffering from ILI by building the sum from k=1 to P, where P is the lagged order of the dependent variable $ILI_{i,t-k}$ that is the number of CDC ILIs in the i-th region in week $t - k$. α_k is the model coefficient and and ϵ is the residual of the model, which is the difference between the observed value of the dependent variable and the predicted value.

Since the relationship between two regions of influenza-affected populations is affected by the relationship between people, an outbreak of influenza in a region with frequent population movements is likely to cause the outbreak of the disease in close regions. To take the movements into account, the weighted CDC regression model (Figure 2.10) was established. Therefore, historical CDC data was used to obtain the correlation coefficient between each region as the weight of the relationship, meaning the greater the correlation, the greater the $w_{i,j}$ value. The parameters β_k (coefficient), $ILI_{i,t-k}$ and $theta_k$ are the same as in model 4. While $w_{i,j}$ represents the weight of the relationship between region i and region j and N represents the number of regions included in the model.

To fit the weights of the model they used a BP Neural Network, its structure is shown in Figure 2.11.

2.5 Influenza prediction through social networks

The diffusion of information through social networks is a hot topic of Big Data research. Information diffusion is a vast research domain which has attracted many researchers from different domains [Gui13]. The spread of a disease among a population has been studied for centuries and several researchers focused on the platform Twitter to do data analysis, including flu trend monitoring and influenza spread prediction [Bu 18].

Nigel Collier et al. [Col11] used SVM algorithm to analyse the epidemic situation by studying the behaviour in the information posted by users on Twitter and compared the results with the ones of the CDC, and found out that it had a very strong correlation.

Lampos, V. et al. [Lam10] observed and tracked the Twitter information published by users in the UK's most popular 49 regions. Using the flu keyword

weighted filtering method, it was found that the influenza episode showed strong linear correlation with the HPA's (Public Health England) ILI data.

In their article, C. Allen et al. [All16] applied techniques from geographic information science (GIS), Data mining and Machine Learning methods to Twitter data for multiscale surveillance of influenza. They collected, filtered and analysed tweets during the 2013-2014 flu season. The results of this procedure have then been compared with national and local flu reports revealing a statistically significant correlation between the data sources.

Retrospective analysis and current estimates are important as they can describe the observed trends. However, further prediction of future flu levels can represent a big leap because such predictions provide realisable insights for health care that can be used for planning, resource allocation, treatments and prevention.[All16]

In contrast to many other approaches, Lee et al. [Lee17] propose a system that not only estimates current flu activity more accurately, but also forecasts future influenza activities a week in advance using aggrefated ILI data by CDC and real-time Twitter data. The results show that their proposed model using MLP with BP algorithm can forecast both current and future influenza activities with high accuracy.

Furthermore, Hu et al. [Hu18] used a Twitter data set and the CDC's influenza-like illness data set to predict ILI in the United States by use of an artificial neural network optimised by the artificial tree (AT) algorithm. They created a model called IAT-BPNN, which stands for improved AT optimising Back-Propagation Neural Network. First they used the basic BPNN for prediction to revise some missing data. Then they performed three different BPNN models to train and test data dynamically. Similar examples of influenza predictions based on social platforms are numerous.
Chen et al. [Xie13] used Facebook, micro-blog, and Instagram as research data to filter textual data for influenza symptoms keywords, to obtain suspected influenza users, and to associate GPS information on Instagram to geographically monitor the flu.

2.6 Influenza prediction with Deep Learning methods

S. Palao et al. [Pal17] give a proposal for epidemic prediction using Deep Learning in their paper. They use a Deep Neural Network which receives training from the past experiences of data they collected from hospitals, government databases and their spread network. Using this training the Deep Neural Network makes predictions for the future, as well as it learns from the real time data. The system after making predictions sends SMS notification to the residents of the area and let the health authorities maintain and manage the predictions using visualisation.

Y. Bu et al. describe in their paper [Bu 18] that it is feasible to predict influenza based on web search data. They used Python web crawler to obtain relevant influenza data from the National Influenza Center's influenza surveillance weekly report and Baidu index. The establishment of support vector regression (SVR), least absolute shrinkage and selection operator (LASSO), Convolutional Neural Network prediction models through Machine Learning, also established the time series model ARMA. Comparing the different models they lead to the result that ARMA model while taking into account also the seasonal characteristics of influenza predicts better results and has a larger generalisation.

Google Flu Trends proposed a method that uses online search queries data to estimate current (real-time) influenza activity. Launched in 2008, GFT used queries from the company's search engine to form prediction models on where the flu was spreading. From 2011 to 2013 GFT failed consistently as it overestimated flu prevalence enormously. The rise and fall of GFT remind us that forecasting an annual event on the basis of one year of data is in a way using only a single data point and thus runs into fundamental time-series problems [Che17].

H. Xue et al. [Xue18] established a MRM also based on GFT and CDC data. They proposed a combination of GFT data and CDC data regression model including seasonal information to receive a better performance.

A method for creating models for predicting peak influenza visits per all health visits has been developed using advanced data mining techniques based on different epidemiological and environmental data by Buczak, Guven et al. in their paper [Buc16].

In their article, N. Ibrahim et al. [Ibr17] describe the use of a Back-Propagation method for the counter measure and prediction analysis of epidemic diseases. The predictive analysis can be determined via machine learning process that promotes the AI in pattern recognition, statistics and features selection. This computer-based learning process is integrated in data mining through classification technique which features disease spread factors that have probably connection between in each other in causing disease outbreaks.

2.7 Markov chain Monte Carlo

A Markov chain is "a stochastic model describing a sequence of possible events in which the probability of each event depends only on the state attained in the previous event" [Uni19b]. For non-infectious diseases, Markov models have been proposed, where latent parameters follow Gaussian Markov random field models.

Markov chain Monte Carlo (MCMC) methodology provides enormous scope for realistic statistical modelling. MCMC methods supply a unifying framework

that helps to analyse many complex problems using generic software. MCMC is essentially Monte Carlo integration using Markov chains.

Bayesian MCMC is a powerful, widely used sampling-based estimation approach. In their paper Li et al. [Li18] describe the application of relatively simple MCMC approaches to data from simulated epidemics that include stochasticity in both transmission and observation, as well as variable generation-interval distributions. They compared model approaches of varying complexity, including an estimation model that matches the simulation model. In addition, they fitted models to simulated epidemic data with multiple sources of heterogeneity, using several different platforms. Using models that include some form of overdispersion in the transmission process can work even better than the full model. Furthermore, they found out that simplifying the models by using continuous rather than discrete latent variables increased efficiency with little effect on the quality of the fits.

2.8 Other approaches

In the article [Car06], the authors a model for simulating the spread of influenza virus infection in the community during a pandemic. They consider not only individual parameters which include the risk of infection and the dynamics of viral shedding according to age, treatment and vaccination status, but also community parameters, in which meetings between individuals are simulated by the use of a complex random graph. Their tool can help to identify the interventions most likely to contain an influenza pandemic.

In their paper, C. Pellat et al. [Pel07] show results of the development of an online computer application allowing analysis of epidemiological time series. It allows to choose between different types of analysis, e.g. retrospective detection and prospective surveillance.

3 Material and methods

In this chapter, details about the implementation, such as the programming language, chosen frameworks and libraries, will be presented (Section 3.3). Since the data sets are essential for the performance of especially ANNs, the used data is described in Section 3.4. Also, the final regression model, as well as details about the NN architecture will be delineated.

3.1 Influenza Surveillence Centres

National influenza centres are institutions for the laboratory surveillance of influenza. They are designated by national ministries of health and formally recognised by WHO for the purpose of participating in the work of the WHO Global Influenza Surveillance and Response System (GISRS) [WHO19].

Furthermore, the most important WHO collaborating centres are the Centres for the Surveillance, Epidemiology and Control of Influenza Centers for Disease Control and Prevention (CDC) in the United States of America [CDC18]. This is where the authors of the presented article [Xue18] in Chapter 2 got their data from.

Moreover, the European Centre for Disease Prevention and Control [EU19], which is an agency for the European Union, plays an important role for the surveillance and prediction of influenza in Europe.

In addition, Sentiweb.fr [Sen19b], which is provided by the Réseau Sentinelles, monitors diseases, influenza is one of them. Réseau Sentinelles is a national surveillance system that enables the collection, analysis, prediction and immediate reproduction of current epidemiological data resulting from the work of general practitioners. It is also part of the French health administration.

3.2 Choice of the models

After intense research of the state-of-the-art models for influenza monitoring and prediction, the decision was made on a Multiple Linear Regression model, that was, inter alia, presented in the article of Xue et al.[Hu18]. In addition, an ANN was created in order to compare the two approaches regarding their prediction results.

3.3 Programming language, frameworks and libraries

3.3.1 Python

Python is a high-level interpreted programming language for many different purposes. Released in 1991 by Guido van Rossum, it gained popularity, is one

Projections of future traffic for major programming languages

Future traffic is predicted with an STL model, along with an 80% prediction interval.
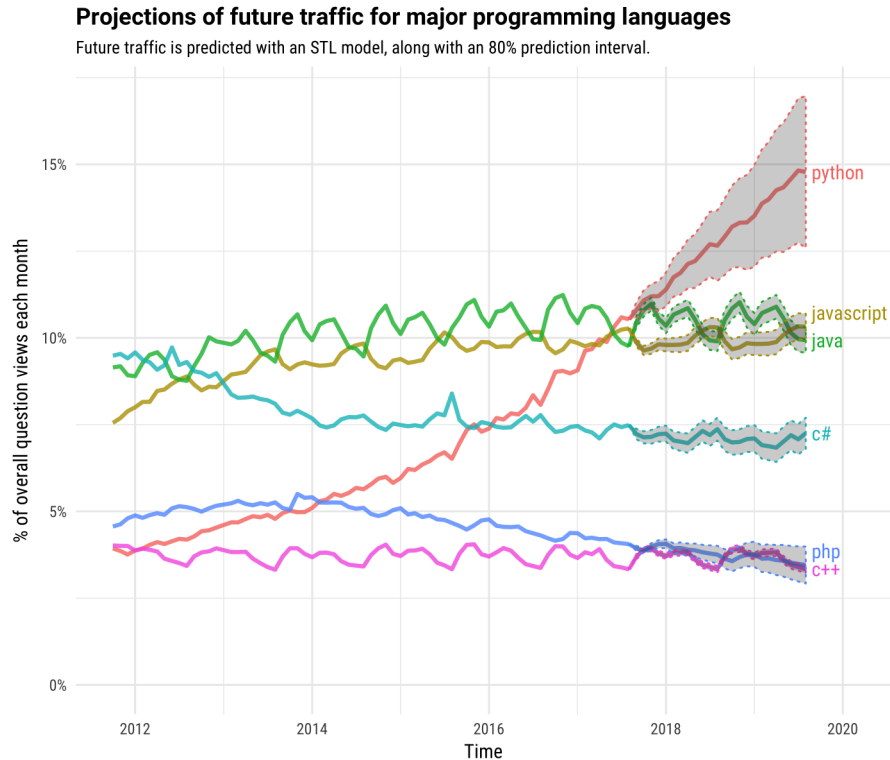

Figure 3.1: Programming language trends
[Ind19]

of the most popular languages for Data Science and has a large community that develops a large number of add-on libraries [Ras16]. With libraries such as NumPy and SciPy, Python is able to compute fast operations on multi-dimensional arrays, which is an important feature for DL purposes. Python is at the moment the most popular programming language with the trend to become the most widely used language in the future two years (as can be seen in Figure 3.1). Since the Python Deep Learning library Keras is currently (as at February 2019) only compatible with Python 2.7-3.6, the latest version, id est Python 3.6 was used for the implementation part of this work.

3.3.2 Pandas

Pandas [pan19] is an open source library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language. It is a NumFOCUS sponsored project that is supposed to ensure the success of development of Pandas as a world-class open-source project, and makes it possible to donate the project. Python has been successful for data munging and preparation for a long time, but less so for data analysis and modeling. Pandas is supposed to fill this gap, enabling to carry out the entire data analysis workflow in Python without having to switch to a more domain specific language like R. In this project DataFrame, a fast and efficient object for data manipulation with integrated indexing, was used. Moreover, a

function for for merging and joining of data sets exists. Furthermore, tools for reading data for csv-files, as well as time-series functionality were used.

3.3.3 Scikit-learn

Scikit-learn is a free ML library for the Python programming language. It features various classification, regression and clustering algorithms and is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy. It is a simple and efficient tool for Data Mining and data analysis. Besides that, it is accessible to everybody and reusable in different contexts as it is open source. Among other features, Scikit-learn provides a function to separate data sets in training and test sets, called "$train\_test\_split$", preprocessing functionality, such as "$MinMaxScaler$" and metrics functions to evaluate the model, such as "$r2\_score$" [sci19].

3.3.4 TensorFlow

TensorFlow is an open source software library for high performance numerical computation and data stream-oriented programming. Its flexible architecture allows easy deployment of computation across a variety of platforms (CPUs, GPUs, TPUs) and from desktops to clusters of servers to mobile and edge devices. It is used within Python programs and is implemented in Python and C++. Originally, TensorFlow was developed by researchers and engineers from the Google Brain team within Google's AI organisation. It comes with strong support for Machine Learning and Deep Learning and the flexible numerical computation core is used across many other scientific domains. [ten19]

3.3.5 Keras

Keras is a high-level Neural Network API, written in Python. Since it is written in Python, it can run on all the major operating systems. Keras is both CPU (Central Processing Unit) and GPU (Graphics Processing Unit) compatible and runs seamlessly on either processing unit. If it is desired to train a network with a small data set, then it will be most likely fine with solely using the CPU. Otherwise it is recommended to use the GPU to run the programs.[Ker18]

Keras is a wrapper that allows you to either use Theano or TensorFlow in the back end. That means, that it is possible to switch between them, depending on your application. Keras also supports Convolutional Networks (CNNs) and Recurrent Neural Networks (RNNs), as well as combinations of the two.[Ker18]

As can be seen in Figure 3.2, Keras was the most widely used framework in articles published by Medium.
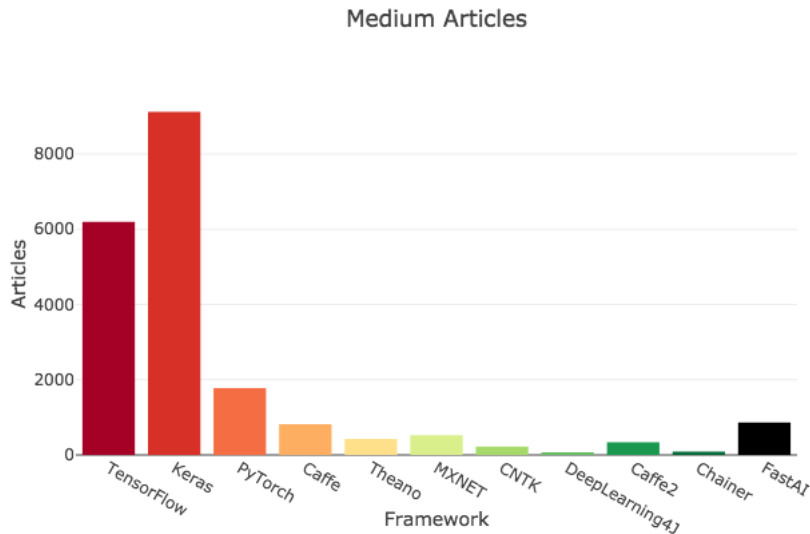Medium is an online publishing platform developed by Evan Williams, and

Figure 3.2: Medium Articles in 2018
[Hal19]

launched in August 2012. The platform is an example of social journalism, having a hybrid collection of amateur and professional people and publications, or exclusive blogs or publishers on Medium, and is regularly regarded as a blog host. [Cor19] Additionally, it is the place for popular Data Science articles and guides. These results might have occurred because Keras is considered to be beginner friendly. Medium is also popular among new DL practitioners and is often used as a forum for tutorials [Hal19].

Keras ($\kappa\varepsilon\rho\alpha\varsigma$) means horn in Greek. It is a cross reference to a literary image from antique Greek and Latin literature. First found in the Odyssey, where dream spirits are divided between those who cheat men with wrong visions, who arrive to Earth through a gate of ivory, and those who announce a future that will come to pass, who arrive through a gate of horn. It is, according to this reference, a play on the words horn/fulfill and ivory/deceive.[Ker18]

3.3.5.1 Guiding principles

Keras follows four main guiding principles. The first is User Friendliness. Since Keras is an API designed for human beings, not machines, it puts user experience front and centre. Keras follows best practices for reducing cognitive load: it offers consistent & simple APIs, it minimises the number of user actions required for common use cases, and it provides clear and actionable feedback upon user error.
The second principle is Modularity. A model is understood as a sequence or a graph of standalone, fully-configurable modules that can be plugged together with as few restrictions as possible. In particular, neural layers, cost functions, optimisers, initialization schemes, activation functions, regularisation schemes

```
week, indicator, inc, inc_low, inc_up, inc100, inc100_low, inc100_up, geo_insee, geo_name
201903,3,11910,8420,15400,210,149,271,44,GRAND EST
201903,3,13137,9275,16999,216,153,279,75,NOUVELLE-AQUITAINE
201903,3,14404,11127,17681,178,138,218,84,AUVERGNE-RHONE-ALPES
201903,3,3090,1786,4394,107,62,152,27,BOURGOGNE-FRANCHE-COMTE
201903,3,3277,1768,4786,96,52,140,53,BRETAGNE
201903,3,5205,3730,6680,197,141,253,24,CENTRE-VAL-DE-LOIRE
201903,3,916,548,1284,273,163,383,94,CORSE
201903,3,32970,25193,40747,269,206,332,11,ILE-DE-FRANCE
201903,3,17646,13524,21768,297,228,366,76,OCCITANIE
201903,3,10301,6530,14072,169,107,231,32,HAUTS-DE-FRANCE
201903,3,2487,1006,3968,73,30,116,28,NORMANDIE
201903,3,4966,1105,8827,129,28,230,52,PAYS-DE-LA-LOIRE
201903,3,13707,8887,18527,269,175,363,93,PROVENCE-ALPES-COTE-D-AZUR
201902,3,6087,3964,8210,107,70,144,44,GRAND EST
201902,3,5340,2800,7880,88,46,130,75,NOUVELLE-AQUITAINE
201902,3,8404,6297,10511,104,78,130,84,AUVERGNE-RHONE-ALPES
201902,3,1398,532,2264,48,18,78,27,BOURGOGNE-FRANCHE-COMTE
201902,3,2170,878,3462,64,26,102,53,BRETAGNE
201902,3,2154,1328,2980,81,50,112,24,CENTRE-VAL-DE-LOIRE
201902,3,293,133,453,87,39,135,94,CORSE
201902,3,20349,14795,25903,166,121,211,11,ILE-DE-FRANCE
201902,3,10032,7068,12996,169,119,219,76,OCCITANIE
201902,3,4905,2847,6963,80,46,114,32,HAUTS-DE-FRANCE
201902,3,2036,890,3182,60,27,93,28,NORMANDIE
201902,3,1912,841,2983,50,22,78,52,PAYS-DE-LA-LOIRE
201902,3,8585,5569,11601,168,109,227,93,PROVENCE-ALPES-COTE-D-AZUR
201901,3,3540,1942,5138,62,34,90,44,GRAND EST
201901,3,2062,430,3694,34,7,61,75,NOUVELLE-AQUITAINE
```

Figure 3.3: Sentiweb ILI data
[Sen19b]

are all standalone modules that can be combined to create new models.

The third principle is Easy Extensibility. New features are simple to add, such as new functions and classes, and existing modules provide extensive examples. To be able to easily create new modules allows for total expressiveness, making Keras suitable for advanced research.

The fourth is the work with Python. There is no need of separate models configuration files in a declarative format. Models are described in Python code, which is compact, easy to debug and allows for effortlessness of expandability[Ker18].

3.4 Implementation

In comparison to logistic regression, which is used when the outcome has only a limited number of possible values and is therefore categorical in nature, the Linear Regression model is used when the response variable is continuous in time. As the number of influenza-infected people is non-categorical, a Multiple Linear Regression Model was chosen. First, an easy MRM was implemented to get familiar with the concept. Therefore, the programming language Python was chosen and the model was tested with some randomly created data in csv-format. As the results were positive, the programme was tested with real data.
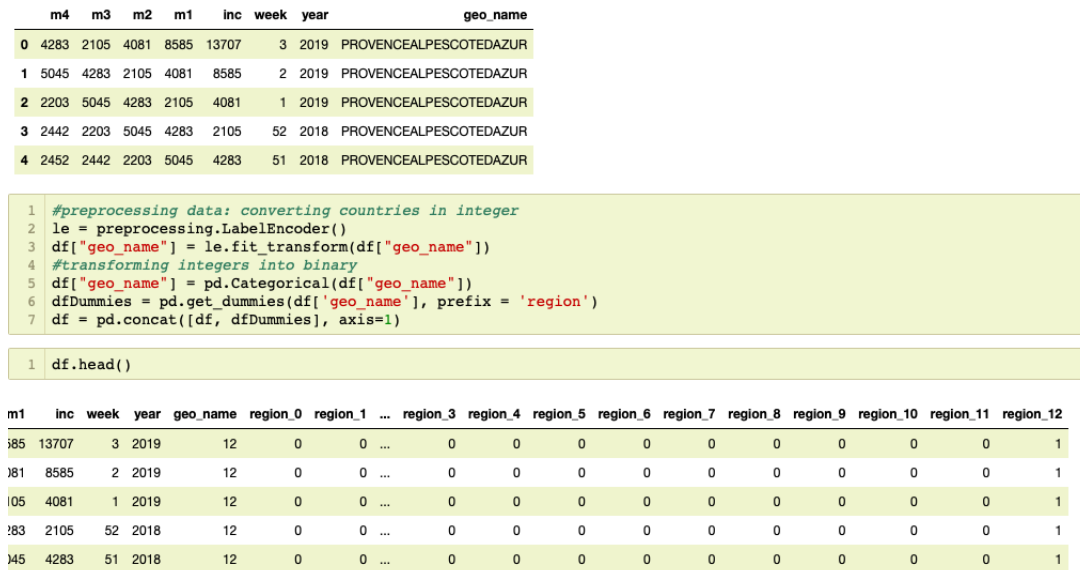
	m4	m3	m2	m1	inc	week	year	geo_name
0	4283	2105	4081	8585	13707	3	2019	PROVENCEALPESCOTEDAZUR
1	5045	4283	2105	4081	8585	2	2019	PROVENCEALPESCOTEDAZUR
2	2203	5045	4283	2105	4081	1	2019	PROVENCEALPESCOTEDAZUR
3	2442	2203	5045	4283	2105	52	2018	PROVENCEALPESCOTEDAZUR
4	2452	2442	2203	5045	4283	51	2018	PROVENCEALPESCOTEDAZUR

```
1  #preprocessing data: converting countries in integer
2  le = preprocessing.LabelEncoder()
3  df["geo_name"] = le.fit_transform(df["geo_name"])
4  #transforming integers into binary
5  df["geo_name"] = pd.Categorical(df["geo_name"])
6  dfDummies = pd.get_dummies(df['geo_name'], prefix = 'region')
7  df = pd.concat([df, dfDummies], axis=1)
```

```
1  df.head()
```

m1	inc	week	year	geo_name	region_0	region_1	...	region_3	region_4	region_5	region_6	region_7	region_8	region_9	region_10	region_11	region_12
585	13707	3	2019	12	0	0	...	0	0	0	0	0	0	0	0	0	1
081	8585	2	2019	12	0	0	...	0	0	0	0	0	0	0	0	0	1
105	4081	1	2019	12	0	0	...	0	0	0	0	0	0	0	0	0	1
283	2105	52	2018	12	0	0	...	0	0	0	0	0	0	0	0	0	1
045	4283	51	2018	12	0	0	...	0	0	0	0	0	0	0	0	0	1

Figure 3.4: Data preprocessing of the regions

3.4.1 Data sets

For the implementation of the two models, the MRM and the NN, different data sets were tried out.

First, the functionality of the model with simple self-created data, containing five input parameters and one output, was tested. After the functionality of the model had been validated, for the task appropriate data was searched.

As Xu et. al used CDC data in their work, it was tried to find similar data sets containing the number of influenza-infected people, the region, the year and the week.

As ECDC did not provide an adequate amount of data fulfilling the requirements, the search had to be continued.

Further research led to the French page of Réseau Sentinelles that places Influenza-like illness data at disposal. Finally, a data set containing the monitored data of all influenza cases since 1985 was found. The data is available for thirteen regions of France, including the week and year and its correlated incidence of influenza cases.

An excerpt of the raw data in csv-format is given in Figure 3.3.

3.4.1.1 Data preprocessing

Due to the fact that some of the Sentiweb data was not numeric, it was necessary to perform a preprocessing to make it comprehensible for the machine. Therefore, the regions were first transformed into categorical integers and after that into binaries, as is shown in Figure 3.4. After having split the data set into input and output variables, the "$train\_test\_split$"-function divided the data set in training and test sets. The whole test and training data was also scaled with the "$scale$"-function of Scikit-learn's preprocessing package.

```
1   #defines "deep" model and its structure
2   model = Sequential()
3   model.add(Dense(16, input_shape=(19,), activation='relu'))
4   model.add(Dense(8, activation='relu'))
5   model.add(Dense(4, activation='relu'))
6   model.add(Dense(2, activation='relu'))
7   model.add(Dense(1, ))
8   optimiser = keras.optimizers.Adam(lr=0.001, beta_1=0.9, beta_2=0.999, epsilon=None, decay=0.0, amsgrad=False)
9
10  #compiles model
11  model.compile(loss='mean_absolute_error', optimizer=optimiser, metrics=['mse', 'mae'])
12
13  # Pass several parameters to 'EarlyStopping' function and assigns it to 'earlystopper'
14  earlystopper = EarlyStopping(monitor='val_loss', min_delta=0, patience=15, verbose=0, mode='auto')
15
16  #Fits model
17  history = model.fit(X_train, y_train, epochs = 100, validation_split = 0.2, shuffle = True, verbose = 0, callbacks
18  history_dict=history.history
```

Figure 3.5: Deep Neural Network Implementation

3.4.2 Definition

The easiest way of creating a model in Keras is by using the sequential API, which lets you stack one layer after another. The problem with the sequential API is that it does not allow models to have multiple inputs or outputs, which are needed for some problems. Nevertheless, the sequential API is a perfect choice for most problems. To create a Neural Network, only a sequential object needs to be created and the "*add*"-function needs to be used to add more layers. An excerpt of Neural Network's architecture is shown in Figure 3.5.

3.4.3 Compilation

Compilation is always required after defining a model. It transforms the simple sequence of layers into a highly efficient series of matrices in a format intended to be executed on your GPU or CPU, depending on how Keras is configured. Compilation requires a number of parameters to be specified, particularly tailored to training the network. Especially, the optimisation algorithm, which is used to train the network and the loss function, which is used to evaluate the network, are minimised by the optimization algorithm.[Bro19]

3.4.4 Fitting and Training

Given that the model's architecture had been constructed, it was ready for training. Training involves, making a prediction based on the current state of the model, calculating how incorrect the prediction is, updating the weights or parameters of the network to minimise this error and, therefore, leading the model to make better predictions. To train a model usually the "fit" method with x, the input data, y, the output data, as well as the number of epochs and the batch size, would be used. It is also passed to a validation set to be able to monitor the loss and different metrics on both sets. Moreover the steps per epoch, which are required when using a generator, are set to the length of the training set divided by the batch size. This process has to be repeated until the model has converged and can no longer learn. There are three key parameters to be chosen for this process: metrics, loss function and optimiser.[Ker18]

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \tilde{y}_i)^2$$

Figure 3.6: Mean Squared Error (MSE)
[med19]

3.4.5 Optimiser

The choice of an optimisation algorithm for a Deep Learning model can have a huge influence on the training time of the network. The Adam optimisation algorithm is an extension to stochastic gradient descent that has seen broader adoption for Deep Learning applications in computer vision and natural language processing. Adam is an adaptive learning rate optimisation algorithm that has been designed specifically for training Deep NNs. Its popularity is growing very fast. First, it calculates an exponentially weighted average of past gradients, and stores it in variables before bias correction and with bias correction. Then it calculates an exponentially weighted average of the squares of the past gradients, and stores it in two other variables before bias correction and with bias correction. Finally, Adam updates parameters in a direction based on combining information.[Bro19]

3.4.6 Metrics

A metric is a function that is used to measure the performance of a model. Metric functions have to be supplied in the metrics parameter. A metric function is similar to a loss function, except that the results from evaluating a metric are not used when training the model. You may use any of the loss functions as a metric function. Sci-kit learn provides different metrics functions for regression, among others, Mean Squared Error (MSE), Mean Absolute Error (MAE) and R-squared score.

3.4.7 Loss function

A loss function is one of the parameters required to compile a Neural Network model. An optimisation problem consists of maximising or minimising a real function by systematically choosing input values from an allowed data set and computing the value of the function. The generalisation of optimisation theory and techniques to other formulations constitutes a large area of applied mathematics. More generally, optimisation includes finding best available values of some objective function given a defined input, including a variety of different types of objective functions. MSE is the most commonly used regression loss function. MSE is the sum of squared distances between the target variable and predicted values, as shown in Figure 3.5 [med19].
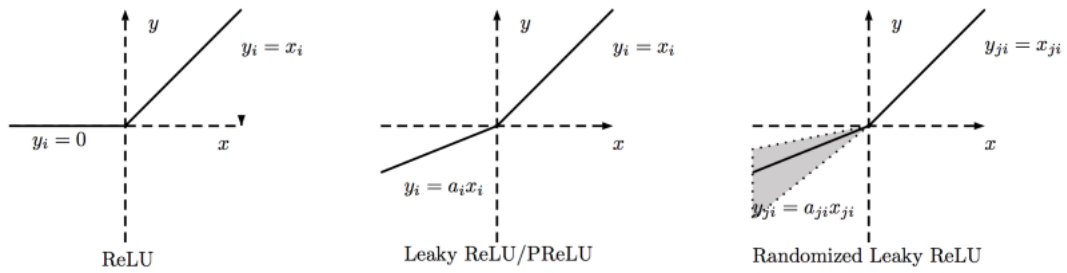
Figure 3.7: ReLU functions
[Tow18]

3.4.8 Activation function

In ANNs, the activation function of a node defines its output, or neuron, given an input or set of inputs. This output is then used as input for the next node and so on until a desired solution to the original problem is found. It maps the resulting values into the desired range such as between 0 to 1 depending upon the choice of activation function. A standard computer chip circuit can be seen as a digital network of activation functions that can be on or off, depending on input. This is similar to the behaviour of the linear perceptron in NNs.

3.4.8.1 The Sigmoid function

One kind of modelling neurons are the so called sigmoid neurons. The difference to the step function, which was explained in Chapter 2, is that their inputs can take any values between 0 and 1 and not only the binary. Like a perceptron, the sigmoid neuron has weights for each inputs and an overall bias b. The output is $\delta(z) = \frac{1}{1+exp(-z)}$.
The sigmoid function has a range between 0 and 1. It is a shaped curve, which is easy to understand and apply, but it has major reasons which have made it fall out of popularity. The first reason is its Vanishing Gradient Problem (VGP). Secondly, the fact that its output is not zero centred makes the gradient updates go too far in different directions and makes optimisation harder. In addition, sigmoids have slow convergence. [Tou]

Due to the VGP simoid and tanh functions should not be used anymore. It causes a lot of problems to train, degreeds the accurancy and performance of a Deep Neural Network.

3.4.8.2 The ReLU function

Nowadays it is common to use the non-linear ReLU activation function $f(x) = max(0, x)$, but only within hidden layers. ANNs train a lot faster with ReLU function than with $tanh$ functions. Nevertheless, ReLU has a significant problem which is that whenever the unit is not active, its gradient is 0. Due to the fact that a gradient-based optimisation algorithm will not adjust the

weights of a unit that never activates initially, it could lead to situations where a unit never activates If the model suffers from dead neurons during training one should use leaky ReLU or Randomised Leaky ReLU, also called "$Maxout$"-function, which are shown in Figure 3.7.

3.4.9 Evaluation

First, it has to be distinguished if the problem is a type of classification or regression. In this case of disease prediction, it is an issue of regression. Therefore, MAE, MSE, as well as R-squared were chosen to evaluate the model.

4 Results

This chapter presents the obtained results of the implemented Multiple Regression model (MRM) and the Artificial Neural Network (ANN), presented in Chapter 3. The results include graphics showing the comparison of actual and predicted output variables in numbers and charts, as well as the metric scores.

4.1 Multiple Linear Regression

First, the MRM model was created and tested with epidemiological SIR data to see if it can perform well. An extract of the data is illustrated in the charts (Figure 4.1 and 4.2), showing the actual and predicted outputs (y_actual and y_pred) of the MRM and the NN. After the preprocessing of the data and the splitting into train and test set analogous to the NN, the final Multiple Regression was performed with the Sentiweb data using Scikit-learn's method "LinearRegression". The resulting scores can be seen in the table below showing the metrics MAE, MSE and R-squared.

4.2 Neural Network for Regression

Analogous to the MRM, the NN was created and tested with epidemiological SIR data to see if it can perform well. An extract of the data is illustrated in the Figure 4.2 showing the actual and predicted outputs (y_actual and y_pred). In order to obtain acceptable good results by running the Artificial Neural Network with surveillance data of Sentiweb, trial and error was indispensable. Therefore, the number of layers and neurons, the learning rate, the optimiser, as well as the batch size and the epochs had to be fitted.
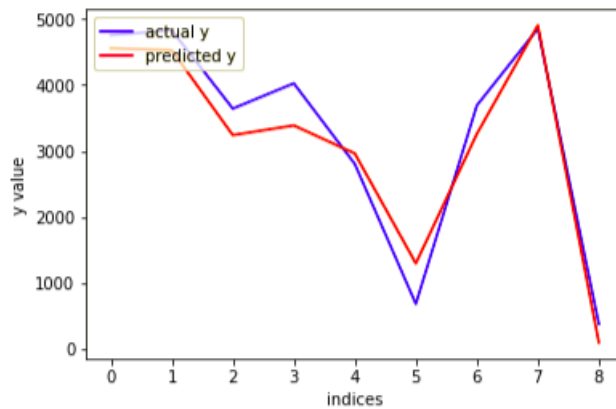

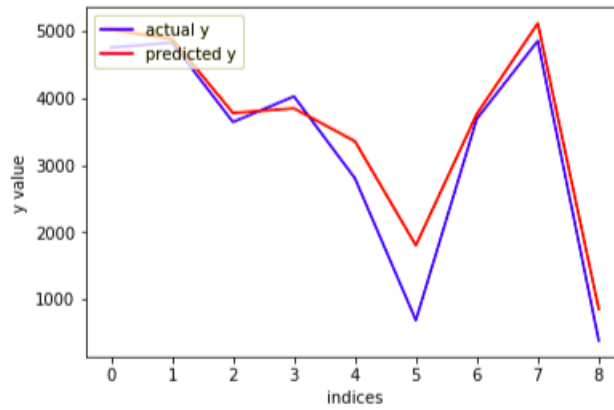Figure 4.1: MRM: Actual and predicted output y

Figure 4.2: NN: Actual and predicted output y

Finally, a five-layer-Neural Network with 16 neurons in the input layer, descending in powers of two was defined. The learning rate was set to 0.0001. Adam was chosen as the optimiser and the epochs parameter was set to 100.

4.3 Comparison of results by means of the metrics

In the table below (Figure 4.3) the metrics Mean Absolute Error (MAE), Mean Squared Error (MSE) and R-squared of the implemented final models are compared. Although the difference between the scores is not enormous, it can be seen that the NN achieved better results in all of the three metrics.

	Multiple Regression	Neural Network
MAE	1957.5644	1558.9771
MSE	23900938.0329	19333329.03875
R-squared	0.7829	0.8244

4.4 Comparison of the predictions

In the table below, an extract of the obtained predictions is shown. y[n] represents the actual number of infected people, while ŷ[n] stands for the predicted values. By regarding only this sample, it can be seen that the predictions of the Neural Network are closer to the actual values than the Multiple Regression model.

	MRM	NN
y[1]	98	98
ŷ[1]	600	411
y[50]	4069	4069
ŷ[50]	6351	4665
y[829]	15321	15321
ŷ[829]	18985	17564
y[1894]	2256	2256
ŷ[1894]	4089	2999

5 Discussion

5.1 Comparison of results

As delineated in the results in Chapter 4, with an adequate data set, both the Multiple Regression model and the Neural Network achieved good results. After having adapted the learning rate, the number of epochs, the batch size and the optimiser, the Neural Network led to even better results.
To be able to better predict the outbreak and propagation of influenza, a Multi-Layer-Neural-Network can be recommended. As several input variables with different units and metrics are needed to run the model, such as the region, the year, the week and the previous incidences of the disease, a preprocessing of the data is very important.
For example, to distinguish between the different seasonal occurrences and to recognise the fluctuations, a Neural Network can perform better than a Multiple Regression model.

Given that input and output often do not correlate linearly, a simple mathematical equation cannot be sufficient. Therefore, it is necessary to consider Neural Networks, as they are also able to include a huge amount of different parameters in their predictions.
Moreover, the learning process of a Neural Network can be more complex than the one of a human brain, and can hence lead to a, for a human being, not so easily comprehensible calculation. Nevertheless, the use of ANNs is indispensable and a major opportunity to take into account several factors, that cannot be regarded by using the existing Multiple Regression Model.

5.2 Ideas to improve influenza prediction

To predict the next occurrence of a seasonal influenza epidemics, it is essential to consider a great amount of training data including several parameters.
This means not only to have the demographic data of one region, one country or one continent, but to monitor as much data as possible in every country of the world.
Therefore, it is inevitable to first install national or regional surveillance centres to monitor and analyse the collected data locally, like the Centers for Disease Control and Prevention already do in the United States of America.
In addition, a continental collaboration between the national centres like the European Centre for Disease Prevention and Control is required.

Furthermore, worldwide cooperation across continents could lead to the formation of a global surveillance centre for influenza and enable prediction of influenza that is analogous to other infectious diseases.
That is certainly not so easy to realise, as it would cause high costs and it is unclear who would have to bear them in a transborder project.

Additionaly, the collection of health data raises a data security issue. Although throughout the EU, common laws exist, such as the General Data Protection Regulation, country-specific disagreements are very likely to occur.

It must not be overlooked, that collecting personal data is a very sensitive issue, even if the data is handled anonymously afterwards. It is therefore mandatory to exactly study the specific regulations before planning the establishment of a global surveillance centre.

5.3 Outlook

If the idea of a global surveillance centre for influenza was established, it would cause a large impact on future research in the field of influenza.

Having a central data set available, that covers the world population and that is accessible to everyone, would also lead to a wider group of people doing research in influenza prediction or disease forecast methods in general. It cannot be denied that the development of algorithms and the use of IT-tools have brought an enormous progress in forecast methods. They are able to deal with a larger amount of data than a human being could ever be capable to calculate with.

Clearly, the digital age entails lots of advantages and progress. However, it is necessary to have the appropriate input for the algorithms, in order to use this possibility properly. To obtain good results with the training of a Neural Network, it is necessary to consider Big Data, because even the best software is useless without analogous data sets.

As described above, this can be achieved with more surveillance centres that collect monitored data. In addition, other sources may be considered for more precise predictions.

Since social networks play an increasingly important role in this day and age, these resources cannot be neglected anymore. As described in Section 2.5, the strategy of using data of a social network platform has already been presented in several articles. However, it is necessary to pursue and deepen this approach.

The establishment of a Big Data set including data related to influenza of all commonly used social networks, would present an even better resource. Combining this proposed data set with the data collected by the global surveillance centre, would represent an adequate amount of data regarding a disease's propagation.

In conclusion, this Bachelor thesis shows that Artificial Neural Networks can improve predictions for the propagation of influenza, but also that the existing simple mathematical Multiple Regression model already produces remarkably good results. Therefore, the amount and parameters of the training data and the desired accuracy can determine if it is necessary to create a well-performing Artificial Neural Network.

However, it cannot be forgotten that predictions are always based on probability and that making a 100% accurate prognosis cannot be aimed for.

The outbreak of a human influenza depends on different factors, including human immune systems that vary form one person to another, pre-existing conditions, climate, seasons and virus mutation. As stated previously in Chapter 3, many factors, inter alia, the architecture of a Neural Network plays an important role in achieving good predictions.

In addition, to make good real-time predictions, it is essential to consider the right data set. A combination of historical data from the global surveillance centres and current social network data could lead to a very precise prediction of influenza.

List of Figures

Bibliography

[All16]   Allen C, Tsou MH, Aslam A, et al. "Applying GIS and machine learning methods to Twitter data for multiscale surveillance of influenza". *PloS one* 11.7 (2016), e0157734.

[Att10]   Attaluri PK, Chen Z, and Lu G. "Integrated use of three machine learning techniques for influenza virus classification". 2010.

[Aun18]   Aunkofer B. "Machine Learning vs Deep Learning – Wo liegt der Unterschied?" 2018.

[Bou14]   Bounekkar A. "Modélisation et analyse des données pour l'aide à la décision". Habilitation à diriger les recherches. Université Lyon 1, Dec. 2014.

[Bro19]   Brownlee J. *Gentle Introduction to the Adam Optimization Algorithm for Deep Learning*. 2019.

[Bu 18]   Bu Y. Bai J.H. CZ. "The Study on China's Flu Prediction Model Based on Web Search Data". *Journal of Data Analysis and Information Processing, 6, 79-92* (2018). DOI: `https://doi.org/10.4236/jdaip.2018.63006`.

[Buc16]   Buczak AL, Baugher B, Guven E, et al. "Prediction of Peaks of Seasonal Influenza in Military Health-Care Data: Supplementary Issue: Big Data Analytics for Health". *Biomedical engineering and computational biology* 7 (2016), BECB–S36277.

[Car06]   Carrat F, Luong J, Lao H, et al. "A'small-world-like'model for comparing interventions aimed at preventing and controlling influenza pandemics". *BMC medicine* 4.1 (2006), p. 26.

[CDC18]   CDC. *About Flu, https://www.cdc.gov/flu/about/index.html*. 2018.

[Che17]   Chen JH and Asch SM. "Machine learning and prediction in medicine — beyond the peak of inflated expectations". *The New England journal of medicine* 376.26 (2017), p. 2507.

[Che05]   Chen WY and Bokka S. "Stochastic modeling of nonlinear epidemiology". *Journal of theoretical biology* 234.4 (2005), pp. 455–470.

[Cob09]   Coburn BJ, Wagner BG, and Blower S. "Modeling influenza epidemics and pandemics: insights into the future of swine flu (H1N1)". *BMC medicine* 7.1 (2009), p. 30.

[Col11]   Collier N, Son NT, and Nguyen NM. "OMG U got flu? Analysis of shared health messages for bio-surveillance". *Journal of biomedical semantics* 2.5 (2011), S9.

[Cor19]   Corporation AM. *Medium*. 2019.

[Dis18]   Disease Modeling I for. *SI and SIS models*. 2018.

[EDU18]   EDUCBA. *Machine Learning vs. Neural Network*. 2018.

[Edu18]   Edureka. *Neural Network Tutorial – Multi Layer Perceptron*. 2018.

[El 17]   El Kihal F, Rachik M, Zakary O, et al. "A multi-regions SEIRS discrete epidemic model with a travel-blocking vicinity optimal control approach on cells". *Int. J. Adv. Appl. Math. Mech* 4.3 (2017), pp. 60–71.

[Ert18]   Ertel W. *Introduction to artificial intelligence*. Springer, 2018.

[EU19]   EU. *European Centre for Disease Prevention and Control*. 2019.

[Fer06]   Ferguson NM, Cummings DA, Fraser C, et al. "Strategies for mitigating an influenza pandemic". *Nature* 442.7101 (2006), p. 448.

[Fle18]   Flexikon D. *Das Medizinlexikon zum Medmachen*. 2018.

[Fre18]   Freiburg SU. *Ausbreitung von Infektionskrankheiten*. 2018.

[Ghe16]   Ghebrehewet Sam MacPherson Peter HA. *Influenza*. 2016.

[Gol88]   Goldberg DE and John H. "Holland. Genetic algorithms and machine learning". *Machine learning* 3.2-3 (1988), pp. 95–99.

[Gon15]   González-Parra G, Villanueva RJ, Ruiz-Baragaño J, et al. "Modelling influenza A (H1N1) 2009 epidemics using a random network in a distributed computing environment". *Acta tropica* 143 (2015), pp. 29–35.

[Goo16]   Goodfellow I, Bengio Y, and Courville A. *Deep Learning*. `http://www.deeplearningbook.org`. MIT Press, 2016.

[Gui13]   Guille A, Hacid H, Favre C, et al. "Information diffusion in online social networks: A survey". *ACM Sigmod Record* 42.2 (2013), pp. 17–28.

[Hal19]   Hale J. *Deep Learning Framework Power Scores 2018*. 2019.

[Hu18]   Hu H, Wang H, Wang F, et al. "Prediction of influenza-like illness based on the improved artificial tree algorithm and artificial neural network". *Scientific reports* 8.1 (2018), p. 4895.

[Huf04]   Hufnagel L, Brockmann D, and Geisel T. "Forecast and control of epidemics in a globalized world". *Proceedings of the National Academy of Sciences* 101.42 (2004), pp. 15124–15129.

[Ibr17]   Ibrahim N, Akhir NSM, and Hassan FH. "Predictive analysis effectiveness in determining the epidemic disease infected area". *AIP Conference Proceedings*. Vol. 1891. AIP Publishing. 2017, p. 020064.

[Ind19]   India U. *Why Python is the most popular language used for Machine Learning*. 2019.

[Ins19]   Institute RK. *Arbeitsgemeinschaft Influenza*. 2019.

[Kar16]   Karpathy A. *CS231n Winter 2016: Lecture 4: Backpropagation, Neural Networks 1*. Jan. 2016.

[Ker18]   Keras. *Keras: The Python Deep Learning library*. 2018.

[Ker32]   Kermack WO and McKendrick AG. "Contributions to the mathematical theory of epidemics. II.—The problem of endemicity". *Proc. R. Soc. Lond. A* 138.834 (1932), pp. 55–83.

[Ker27]   Kermark M and Mckendrick A. "Contributions to the mathematical theory of epidemics. Part I". *Proc. r. soc. a* 115.5 (1927), pp. 700–721.

[Kri17]   Kriesel D. *Neural Networks*. 2017.

[Lam10]   Lampos V, De Bie T, and Cristianini N. "Flu detector-tracking epidemics on Twitter". *Joint European conference on machine learning and knowledge discovery in databases*. Springer. 2010, pp. 599–602.

[Lee17]   Lee K, Agrawal A, and Choudhary A. "Forecasting Influenza Levels Using Real-Time Social Media Streams". *Healthcare Informatics (ICHI), 2017 IEEE International Conference on*. IEEE. 2017, pp. 409–414.

[Li17]   Li C, Ma J, Guo X, et al. "DeepCas: An end-to-end predictor of information cascades". *Proceedings of the 26th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee. 2017, pp. 577–586.

[Li14]   Li D, Xu Z, Li S, et al. "A survey on information diffusion in online social networks". *Chin. J. Comput* 254 (2014), pp. 31–36.

[Li18]   Li M, Dushoff J, and Bolker BM. "Fitting mechanistic epidemic models to data: a comparison of simple Markov chain Monte Carlo approaches". *Statistical methods in medical research* 27.7 (2018), pp. 1956–1967.

[Lia18]   Liang F, Guan P, Wu W, et al. "Forecasting influenza epidemics by integrating internet search queries and traditional surveillance data with the support vector machine regression model in Liaoning, from 2011 to 2015". *PeerJ* 6 (2018), e5134.

[Mar15]   Martcheva M. *An introduction to mathematical epidemiology*. Vol. 61. Springer, 2015.

[Mar17]   Martineau AR, Jolliffe DA, Hooper RL, et al. "Vitamin D supplementation to prevent acute respiratory tract infections: systematic review and meta-analysis of individual participant data". *bmj* 356 (2017), p. i6583.

[med19]   medium. *Machine learning: an introduction to mean squared error and regression lines*. 2019.

[Mil06]   Miller SJ. "The method of least squares". *Mathematics Department Brown University* (2006), pp. 1–7.

[Mit98]   Mitchell M. *An introduction to genetic algorithms*. MIT press, 1998.

[Nie15]   Nielsen MA. *Neural networks and deep learning*. Vol. 25. Determination press USA, 2015.

[Pal17]   Palao S. Shahasane A. DS. "A Proposal for Epidemic Prediction using Deep Learning". *IJERCSE* (2017). DOI: `01.1617/vol4/iss3/pid01563`.

[pan19]   pandas.pydata.org. *Pandas*. 2019.

[Pap18]   Paperspace H. *Intro to optimization in deep learning: Gradient Descent*. 2018.

[Pat17]   Patterson J and Gibson A. *Deep Learning: A Practitioner's Approach*. O'Reilly Media, Inc., 2017.

[Pel07]   Pelat C, Boëlle PY, Cowling BJ, et al. "Online detection and quantification of epidemics". *BMC Medical Informatics and Decision Making* 7.1 (2007), p. 29.

[Quo18]   Quora. *What is the differences between artificial neural network and biological neural network?* 2018.

[Ras16]   Raschka S, Julian D, and Hearty J. *Python: deeper insights into machine learning*. Packt Publishing Ltd, 2016.

[Rud16]   Ruder S. "An overview of gradient descent optimization algorithms". *arXiv preprint arXiv:1609.04747* (2016).

[Rum85]   Rumelhart DE, Hinton GE, and Williams RJ. *Learning internal representations by error propagation*. Tech. rep. California Univ San Diego La Jolla Inst for Cognitive Science, 1985.

[Rva68]   Rvachev L. "Computer modeling experiment on large-scale epidemic". *Dokl USSR Acad Sci*. Vol. 2. 1968, pp. 294–296.

[Rva85]   Rvachev LA and Longini Jr IM. "A mathematical model for the global spread of influenza". *Mathematical biosciences* 75.1 (1985), pp. 3–22.

[Sar05]   Saramäki J and Kaski K. "Modelling development of epidemics with dynamic small-world networks". *Journal of Theoretical Biology* 234.3 (2005), pp. 413–421.

[Sch15]   Schmidhuber J. "Deep learning in neural networks: An overview". *Neural networks* 61 (2015), pp. 85–117.

[sci19]   scikit-learn. *scikit-learn: Machine Learning in Python*. 2019.

[Sen19a]   Sentinelles EGR. *GrippeNet.fr*. 2019.

[Sen19b]   Sentinelles R. *sentiweb.fr*. 2019.

[Sta18]   Statistics - Columbia University D of. *Introduction to epidemic modelling*. 2018.

[ten19]   tensorflow.org. *Tensorflow*. 2019.

[Tie18]   Tiedemann M. "KI, künstliche neuronale Netze, Machine Learning, Deep Learning: Wir bringen Licht in die Begriffe rund um das Thema „Künstliche Intelligenz"." 2018.

[Tou]   Toulouse U. *Neural Networks and Introduction to DL*.

[Tow18]   TowardsDataScience. *Activation functions and its types*. 2018.

[Uni19a]   University C. *Cambridge Dictionary*. 2019.

[Uni19b]   University O. *Oxford Dictionaries*. 2019.

[Wal19]   Walczak S. "Artificial neural networks". *Advanced Methodologies and Technologies in Artificial Intelligence, Computer Simulation, and Human-Computer Interaction*. IGI Global, 2019, pp. 40–53.

[WHO18]   WHO. *Definition epidemiology*. 2018.

[WHO19]   WHO. *Global Influenza Surveillance and Response System (GISRS)*. 2019.

[Xie13]   Xie Y, Chen Z, Cheng Y, et al. "Detecting and tracking disease outbreaks by mining social media data". *Dimensions* 17.16 (2013), pp. 16–70.

[Xue18]   Xue H, Bai Y, Hu H, et al. "Influenza activity surveillance based on multiple regression model and artificial neural network". *IEEE Access* 6 (2018), pp. 563–575.

Statutory Declaration

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

Heilbronn, February 20, 2019

(place, date)   (signature)