

# Handschriftherkenning voor doktersbriefjes

## Onderzoeksvoorstel Bachelorproef 2019-2020

Helmi Dayyeh<sup>1</sup>

### Samenvatting

Met de opkomst van **artificiële intelligentie** worden alledaagse taken al vaak geautomatiseerd. Vandaag de dag hebben bedrijven de tijd en het budget niet om talloze mensen aan te werven om terugkerende manuele taken uit te voeren. Bovendien zijn mensen gekend om fouten te maken en kan dit een negatieve invloed op de workflow hebben.

Deze analyse is gebaseerd op een recent onderzoek uitgevoerd door Vectr.Consulting in samenwerking met de Christelijke mutualiteit. De probleemstelling bij dit onderzoek bleek te zijn dat de handgeschreven overlijdensakten niet leesbaar waren. Dit heeft men dan doormiddel van een applicatie die gebruik maakt van het Tensorflow framework met een getraind neurale netwerk opgelost, waarbij in **47 %** van de gevallen het handgeschrift correct voorspeld werd.

De vraag en hetgeen dit onderzoek zal gaan vaststellen is of via deze handtering ook het handgeschrift op doktersbriefjes kan herkend worden.

In de literatuurstudie zal tot in detail de opzet en de statistieken van het neurale netwerk beschreven staan. Het grootste obstakel zal dan ook uiteraard de data zijn die verzameld zal moeten worden van een derde partij.

Het praktische gedeelte van dit onderzoek zal gebruik maken een app die in de achtergrond **azure cognitive services** gebruikt om het model te trainen en predicties te maken.

Door dit onderzoek uit te voeren met een praktisch voorbeeld zal de christelijke mutualiteit een meerwaarde zien in de overstap naar **artificiële intelligentie** binnen hun organisatie.

### Sleutelwoorden

Onderzoeksdomein. AI, Machine learning — Azure Cognitive Services

### Co-promotor

Wouter Baetens

## Inhoudsopgave

- 1 **Introductie**
- 2 **State-of-the-art**
- 3 **Methodologie**
- 4 **Verwachte resultaten**
- 5 **Verwachte conclusies**
- 6 **Referenties**

### 1. Introductie

**Probleemstelling:** Handgeschrift op doktersbriefjes is in sommige gevallen onleesbaar voor de christelijke Mutualiteit. Het is nu ook niet wetenschappelijk bewezen dat dokters het slechtste handgeschrift hebben, maar het wordt wel sociaal vastgesteld.

**Motivatie:** Het gebied van Artificiële intelligentie en machine learning spreekt toekomst. Bijna ieder toestel maakt al op een manier gebruik van AI. Door de opkomst van IoT genereren miljoenen gebruikers dagelijks data die gebruikt wordt om de accuraatheid van AI geïntegreerde systemen te vergroten.

We kunnen hier dan ook uit vaststellen dat in de nabije toekomst het gebruikt van AI alleen maar effectiever zal

worden. Het is dus belangrijk dat we hier zo snel mogelijk op inspelen en een domein onderzoeken waar we dit dan ook kunnen gebruiken.

**Doelstelling:** De werkelijke onderzoeksvraag is dus of de workflow efficiëntie van de christelijke Mutualiteit verbeterd kan worden doormiddel van het integreren van handschriftherkenning. Dit zal natuurlijk in meerdere stappen uitgevoerd worden. Door het verzamelen van voorgaande data en het trainen van het neurale netwerk wordt verwacht dat dit model het handgeschrift op doktersbriefjes kan herkennen/ parsen met een efficiëntie van +- 47 %. Dit zal uiteraard naarmate de grote van de dataset verbeteren.

### 2. State-of-the-art

Zoals vooraf al aan bod kwam is deze onderzoeksvraag gebaseerd op een recent onderzoek dat uitgevoerd werd door Vectr.Consulting in samenwerking met de Christelijke Mutualiteit. In dit onderzoek werd een model opgezet en getraind om doktershandgeschrift te herkennen op doodsakten. Verder, baseerde het praktische gedeelte van dit uitgevoerde onderzoek zich op het open source Tensorflow framework. Dit onderzoek zal vooral gebaseerd zijn rond de Microsoft Azure stack, meer bepaald de afdeling Cognitive Services. Microsoft hanteert hiervoor hun krachtig cloud networking

solution om op basis van vooraf getrainde modellen real-time data te bezorgen aan de consument. Dit framework is vooral nog in de early development stages en zal naarmate het gebruik alleen maar verbeteren.

### 3. Methodologie

In zeker zin zal er een observatieonderzoek plaats vinden. De dataset gaat bestaan uit vooraf geschreven doktersbriefjes waarbij het computer vision gedeelte van azure cognitive services de data uit elke image extraheert en analyseert. Door deze methodologie zal de eindgebruiker de mogelijkheid hebben de geëxtraheerde data een rating te geven op basis van accuraatheid van foto naar tekst extractie. Het model zal hierdoor alleen maar bijleren en zichzelf aanpassen naar toekomstige data.

### 4. Verwachte resultaten

Na het trainen en modelleren van de data zal het handschrift op doktersattesten met een nauwkeurigheid van omtrent 47 % voorspeld kunnen worden. Data zal gestructureerd worden in functie van de kwaliteit en hoe zeker het model is van zijn voorspelling. Om dit percentage te bereiken zal het model enkel getraind worden met data die aan een bepaalde kwaliteitsstandaard voldoet en zal na elke voorspelling de eindgebruiker geprompt worden om de kwaliteit van extractie te raten. Medische termen zullen hierbij een zekere rol spelen, aangezien deze niet altijd de makkelijkste zijn om te herkennen.

### 5. Verwachte conclusies

Wanneer de bevindingen concorderen met hetgeen verwacht wordt, wordt er gesteld dat de Christelijke Mutualiteit een meerwaarde in workflow en resources zal bekomen, in de hoop dat meerdere taken door artificiële intelligentie geautomatiseerd zullen worden. Dit is natuurlijk een oplossing die op lange termijn zichzelf zal verbeteren tot op het punt dat de accuraatheid van de voorspelde tekst naar een 80-90 % convergeert.

### 6. Referenties

<https://vectr.consulting/news/deciphering-handwriting>  
<https://vectr.consulting/news/smart-scanner/artificial-intelligence>  
<https://azure.microsoft.com/en-us/services/cognitive-services/computer-vision/>  
<https://towardsdatascience.com/build-a-handwritten-text-recognition-system-using-tensorflow-2326a3487cd5>