# Deep learning for protein model quality assessment – a more startling title should not be impossible to come up with?

Contents

Introduction 1

Background 2

Aim 2

Ethical statement 2

Process 2

Theory 3

Structure models/ MQA 3

An introduction to deep learning 3

Methods 4

Results 4

Discussion 5

Conclusions 6

Future prospects 6

References 6

Rapporten ska:

* innehålla en redogörelse för projektets problemställning och syfte samt ge korrekt vetenskaplig, teknisk och metodologisk bakgrundsinformation baserat på relevanta källor,
* innehålla en beskrivning av erhållna resultat och analys av dessa kopplat till befintliga teorier,
* visa teoretisk förståelse, etisk medvetenhet samt insikt om samhälleliga aspekter.
* en systematisk beskrivning av hur man kommer fram till slutsatser och resultat (se även avsnitten markerade med asterisk (\*) senare), innefattande
  + en detaljerad tidsmässig och metodologisk beskrivning av projektets ursprungsplanering,
  + en kronologisk redogörelse för och analys av projektets framskridande, samt en beskrivning av de beslut som fattats med denna analys som bakgrund.

## Introduction

INTRODUKTIONEN SKA I SIN HELHET INNEHÅLLA:

* Beskrivning av nuvarande kunskapsläge. = Vad kan man om deep learning? Börja med historia och gå framåt.
* Beskrivning av nuvarande teknisk nivå. Hur bra är protein mqa’sen som finns nu?
* Beskrivning av studiens syfte, exempelvis efterfrågad ny kunskap, en mer utvecklad produkt eller en ny/förbättrad metod *genom en tydligt formulerad frågeställning.*
* Beskrivning av hur den frågeställning som beskrivs i syftet tidigare har studerats/ angripits och hur den är tänkt att angripas i den aktuella studien.
* Beskrivning av relevanta metoder på ett övergripande sätt: dels de tidigare använda metoder som gett viktig information, dels de metoder som inte använts tidigare men som nu är intressanta att använda. OBS! Detaljerad beskrivning av de i studien använda metoderna ska ske i metoddelen.
* Information om eventuella avgränsningar.
* En reflektion om eventuella etiska frågeställningar som är kopplade till detta kunskaps-/teknikområde.
* En kortfattad beskrivning av den samhälleliga relevansen för detta kunskaps-/teknik- område och hur det specifika projektet kan ge ett bidrag, exempelvis genom att vara relaterat till industriell produktion, miljöarbete, forensisk verksamhet, läkemedels- utveckling eller sjukdomsdiagnostik.

Tänker mig en inledning likt planeringsrapporten!

Protein folding? 3D struktur/function

Varför det är viktigt att veta hur bra strukturmodeller är?

MQA

Deep learning /machine learning

What? How? Why important? Läget idag. Problems?

”The key aspect of deep learning is that these layers of features are not designed by human engineers: they are learned from data using a general-purpose learning procedure.” [2]

### Background

Biological background (Protein structure & protein folding)

Protein structure prediction

Deep learning

Model Quality Assessment

* + 3DCNN

### Aim

Goals? Research questions?

* Better representations/ more suitable features
* Optimization
* Automatic feature extraction?? – är det något jag kan titta på?

Avgränsningar

### Ethical statement

### Process

Kan med fördel läggas i ett appendix. Gör man så är det viktigt att man tydligt hänvisar till Appendix i den inledande delen av rapporten. Utgå från den planeringsrapport som skrevs i början av examensarbetet.

Beskriv planeringen av hur arbetet är tänkt att läggas upp och hur arbetets framåtskridande ska följas upp. Lämpligt är att börja med att upprätta en tidsplan, exempelvis i form av ett GANTT- schema. Börja avsnittet i rapporten med en introducerande mening som säger att det vid planeringen av arbetet upprättades en tidsplan för dess genomförande. Fördelarna med en tidsplan är flera: Genom att ordna alla tänkta aktiviteter/delprojekt i en kronologisk ordning och genom beräkning av tidsåtgång för var och en av dessa blir det möjligt att kontinuerligt och system- atiskt analysera och värdera hur projektet utvecklas.

**1. Tidsplan**

Den kronologiska tidsplanen skall innehålla en noggrann beskrivning av arbetsmoment och beräknad tidsåtgång. TÄNK: Vad är målet och vilka aktiviteter krävs för att nå målet? Planen ska skrivas med flera konkreta delmål, för att ge möjlighet till kontinuerlig och systematisk uppföljning enligt en dokumenterad plan.

**2. Plan för systematisk uppföljning**

Innehåller a) plan och mål för nästa aktivitet, b) genomförande av aktiviteten, c) analys av resultatet i förhållande till plan och mål för aktiviteten, d) gå till nästa planerade aktivitet eller planera om, e) gör om punkt a-d tills arbetet är färdigt.

## Theory

### Structure models/ MQA

Hur applicerbart med deep learning på detta?

### An introduction to deep learning

A subfield of machine learning.

supervised

#### Convolutional Neural Networks

* Se LeCun! Asbra!
* There are four key ideas behind ConvNets that take advantage of the properties of natural signals: local connections, shared weights, pooling and the use of many layers.

Architecture

Layers / Activations

Training - features

Fitting & overfitting

“Effects of model structure on uncertainty”

## Methods

Datasets / Training / testing

**1. Metodteori**

I de flesta fall är det motiverat att beskriva de teoretiska principerna för (någon eller några av) de använda metoderna. Illustera gärna med figurer och kommentera metodernas eventuella begränsningar (om de är viktiga).

**2. Modeller**

Om matematiska modeller har använts i arbetet, ska dessa beskrivas här. Är det lämpligt att genomföra en statistisk bearbetning av resultatet ska även den statistiska metoden beskriv- as.

**3. Experimental**

Här anges de material, exakta procedurer och betingelser som användes under den genom- förda studien. Redogörelsen ska vara kortfattad men detaljerad och skrivas så att läsaren ska kunna upprepa försöken. Det kan vara lämpligt att dela upp det här avsnittet efter metod. Om avancerade instrument använts för studien ska varumärke och modell framgå och om specialkemikalier krävts ska det framgå vilket företag som levererat dem.

## Results

Här presenteras alla typer av resultat som framkommer under arbetets gång. Sektionen skall innehålla två delar: slutresultat och processanalys\* (om inte det senare istället läggs i Appen- dix men då ska man hänvisa dit).

**1. Slutresultat**

Här beskrivs i text, tabeller och figurer de resultat som har genererat ny kunskap eller har givit upphov till en ny produkt. En produkt har här en vid betydelse, till exempel en kemisk analysmetod, en organisk molekyl med vissa egenskaper, ett enzym med ny kata- lytisk aktivitet, en sensor som registrerar halten av något eller en förbättrad produktions- strategi. Sträva efter att varje figur och tabell i stort sett ska vara självförklarande, d v s att tabell- och figurtexten ger så tydlig information som möjligt. Var noga med att i texten hänvisa till tabeller och figurer på ett korrekt sätt. Finns underlag till en statistisk analys av mätdata ska detta genomföras och redovisas.

**2. Processanalys\***

Denna analys av processen är tydligt beroende av planeringsrapporten och den tidsplan som beskrivs i sektionen ”Process”.

Här skall den genomförda processen beskrivas i kronologisk ordning. Den skall beskrivas stegvis med fokus på de planerade delmålen. Ett typiskt upplägg kan vara:

**Steg 1** Visa ett intermediärt resultat: beskriv kortfattat den analys som gjordes, och ange vilket beslut som togs inför nästa steg. De alternativ som finns är 1) gå vidare enligt ur- sprunglig plan eller 2) förändra planen. Valdes det senare måste förändringen beskrivas på ett tydligt sätt.

**Steg 2** Som ovan, osv

## Discussion

Diskussion kan delas upp i flera delar och en möjlig uppdelning beskrivs nedan.

1. **Analys av de viktigaste resultaten**Vilken ny kunskap har genererats från resultaten? Har alla resultat kommenterats? Har den studie som genomförts gett underlag till att lösa uppgiften? Om endast en del av målen har uppnåtts, vad saknas och hur kan man gå vidare? Är resultaten relevanta och finns några begränsningar att ta hänsyn till? Är resultaten statistiskt säkerställda eller finns mätosäkerheter? Finns några alternativa metoder som skulle kunna gett mer precisa resultat? Finns några svagheter eller felkällor? OBS! Var noga med att åter- knyta till syftet!
2. **Övergripande analys av processen**\*  
   Hur väl lyckades den ursprungliga planeringen? Vilka viktiga förändringar genom- fördes? Hur effektiv var processen för att nå de mål som önskades? Denna del kan mynna ut i en rekommendation om hur en effektiv process bör se ut.

iii) **Resultatets samhälleliga konsekvenser i vid bemärkelse**Finns det någon samhällsnytta med studiens resultat? Oavsett om arbetet haft fokus på att utveckla en process eller en metod, utveckla en produkt eller generera grundläggande ny kunskap så bör diskussionen innehålla detta. Det kan vara resultat som ger allmänheten en ökad kunskap om naturliga fenomen eller nya förutsättningar för dagens snabba högteknologiska utveckling, det kan röra sig om utveckling eller kontroll av en produkt eller en process inom industrin alternativt kan resultatet ha betydelse för eko- logisk eller social hållbar utveckling som verktyg för diagnoser inom sjukvården, foren- siska analyser och så vidare.

iv) **Resultatets etiska konsekvenser i vid bemärkelse**Finns det några etiska överväganden som bör diskuteras? Finns det några säkerhets- eller miljöföreskrifter som är viktiga att ta hänsyn till? Kan det resultat som kommit fram i studien missbrukas och användas i ett annat syfte? Har personer som eventuellt lämnat prover som ligger till grund för resultatet samtyckt till provernas användande? Är djurförsök relevanta för studiens genomförande eller finns andra alternativa modeller? Vilka värden ska vi slå vakt om – finns det någon risk för miljön eller någons hälsa? Eventuella andra frågeställningar?

Avrunda med underrubriken **Framtida utvecklingsmöjligheter (Future perspectives)** där en intressant utveckling av projektet diskuteras. Man kan även lägga till rubriken **Sammanfatt- ning (Conclusions)** där de absolut viktigaste resultaten listas.

## Conclusions

### Future prospects

## References

[1] D. Kihara, H. Chen, and Y. Yang, ‘Quality Assessment of Protein Structure Models’, *Curr. Protein Pept. Sci.*, vol. 10, no. 3, pp. 216–228, Jun. 2009.

[2] Y. LeCun, Y. Bengio, and G. Hinton, ‘Deep learning’, *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.