

UNIVERSITY OF ZAGREB  
FACULTY OF ELECTRICAL ENGINEERING AND COMPUTING

MASTER THESIS num. 1572

# **End-to-End Deep Learning Model for Base Calling of MinION Nanopore Reads**

Neven Miculinić

Zagreb, June 2018.

*Umjesto ove stranice umetnite izvornik Vašeg rada.*  
*Kako biste uklonili ovu stranicu, obrišite naredbu \izvornik.*

*I would like to thank my mentor, Mile Šikić, for his patient guidance, encouragement and advice provided over the years.*

*I would also like to thank my family and friends for their continuous support.*

*In the end, honorable mentions go to Marko Ratković for his help with this thesis.*

# CONTENTS

<b>1. Introduction</b>	<b>1</b>
<b>2. Conclusion</b>	<b>2</b>
<b>Bibliography</b>	<b>3</b>

# LIST OF FIGURES

# LIST OF TABLES

# 1. Introduction

I wanna cite somebody(Oetiket et al.)

Thesis introduction.

## **2. Conclusion**

Conclusion.



# BIBLIOGRAPHY

T. Oetiket, H. Partl, Hyna, and E. Schlegl. *The not-so-short introduction to LaTeX*. URL <http://www.ctan.org/tex-archive/info/lshort/english/lshort.pdf>.

## **End-to-End Deep Learning Model for Base Calling of MinION Nanopore Reads**

### **Abstract**

In the MinION device, single-stranded DNA fragments move through nanopores, which causes drops in the electric current. The electric current is measured at each pore several thousand times per second. Each event is described by the mean and variance of the current and by event duration. This sequence of events is then translated into a DNA sequence by a base caller. Develop a base-caller for MinION nanopore sequencing platform using a deep learning architecture such as convolutional neural networks and recurrent neural networks. Instead of events, use current waveform at the input. Compare the accuracy with the state-of-the-art basecallers. For testing purposes use publicly, available datasets and Graphmap or Minimap 2 tools for aligning called reads on reference genomes. Implement method using TensorFlow or similar library. The code should be documented and hosted on a publicly available Github repository.

**Keywords:** base calling, Oxford Nanopore Technologies, MinION, deep learning, seq2seq, convolutional neural network, residual network, CTC loss

### **S kraja na kraj model dubokog učenja za određivanje očitanih baza dobivenih uređajem za sekvenciranje MinION**

#### **Sažetak**

Unutar uređaja MinION, fragmenti jednostruke DNA prolaze kroz nanopore, što uzrokuje promjene u električnoj struji. Struja proizvedena na svakoj nanopori mjeri se nekoliko tisuća puta u sekundi. Svaki događaj opisan je srednjom vrijednosti i varijancom struje te svojim trajanjem. Postupak kojim se takav slijed događaja prevodi u niz nukleotida naziva se određivanje očitanih baza. Razviti alat za prozivanje baza za uređaj za sekvenciranje MinION koristeći modele dubokog učenje kao što su konvolucijske i povratne neuronske mreže. Umjesto događaja na ulazu koristi valni oblik struje. Usporediti dobivenu točnost s postojećim rješenjima. U svrhu testiranja koristiti javno dostupne skupove podataka i alate GraphMap ili Minimap 2 za poravnanje očitavanja na referentni genom. Alat implementirati koristeći programsku biblioteku TensorFlow (ili neku sličnu). Programski kod treba biti dokumentiran i javno dostupan preko repozitorija GitHub. **Ključne riječi:** određivanje

baza, Oxford Nanopore Technologies, MinION, duboko učenje, prevođenje, konvolucijske neuronske mreže, rezidualne mreže, CTC gubitak