



**Vilniaus
universitetas**

Doktorantas:
Brendonas Stakauskas
2022-2026

Darbo vadovas:
Dr. Virginijus Marcinkevičius

2024/2025 metai
III metai, II pusmetis



**Giliais neuroniniais tinklais grįstų
mašininio mokymo metodų taikymas
viruso mutacijų trajektorijai
prognozuoti**



Studijų plano vykdymas

Studijų planas, vykdymo suvestinė

Vilniaus
universitetas

Studijų metai	Egzaminai		Dalyvavimas konferencijose				Publikacijos				
			Nacionalinėse		Tarptautinėse		Su citavimo rodikliu		Be citavimo rodiklio		Buklė
	Planas	Įvykdyta	Planas	Įvykdyta	Planas	Įvykdyta	Planas	Įvykdyta	Planas	Įvykdyta	
I (2022/2023)	2	2	1	0	0	1	0	0	1	1 (konf. darbų medžiagoje)	
II (2023/2024)	2	2	1	0	0	1	0	0	1	0	
III (2024/2025)	0	0	0	1	1	0	1	0*	0	1 (konf.darbų medžiagoje)	*Įteikta žurnalui Evolutionary Bioinformatics
IV (2025/2026)	0	0	0	0	1	0	1	0	0	0	
Iš viso:	4	4	2	1	2	2	2	0	2	1	

Ataskaitinio pusmečio darbo planas ir jo įvykdymas

Dalyvavimas konferencijose 2024/2025 (2 pusmetis)		
Planas	Įvykdyta	Konferencijos tipas
34th International Conference on Artificial Neural Networks, 2025 Rugsėjo 9-12, Kaunas, Lietuva	Poster B. Stakauskas "Dimensionality Reduction of Protein Language Model Embeddings for Viral Clustering"	Tarptautinė

Ataskaitinio pusmečio darbo planas ir jo įvykdymas

Publikacijos 2024/2025 (2 pusmetis)			
Planas	Įvykdyta	Konferencijos tipas	Publikacijos tipas
Evolutionary Bioinformatics	Evaluating Protein Language Model Embeddings for Viral Clade Assignment	Įteikta: 2025-09-23	CA WoS su Impact Factor (1.5)

Visų mokslinių tyrimų ir disertacijos rengimo etapai

1. Mokslinių tyrimų disertacijos tema apžvalga ir analizė
2. Mokslinio tyrimo vykdymas
 1. Tyrimo metodikos sudarymas
 2. Teorinis tyrimas
 3. Empirinis tyrimas
 4. Gautų duomenų analizė, apibendrinimas, išvadų parengimas
3. Atskirų daktaro disertacijos dalių (tyrimo metodikos, rezultatų, ginamų teiginių, išvadų, ir kt.) parengimas
4. Daktaro disertacijos parengimas ir svarstymas padalinyje
5. Daktaro disertacijos gynimas

Trumpas per pusmetį gautų mokslinių rezultatų pristatymas

Įterpinių palyginimas

- Naudojami ESM-2, ProtBert, ProtT5 ir CARP modelių įterpiniai
- Sumažinama duomenų dimensija
- Naudojamas klasterizavimo metodas HDBSCAN sudarant klasterius
- Klasteriai palyginami su priskirtais kladais
- Rezultatai validuojami su naujesniais gripo variantais
- Tikrinamas modelių ir rezultatų stabilumas

Paieškos erdvė

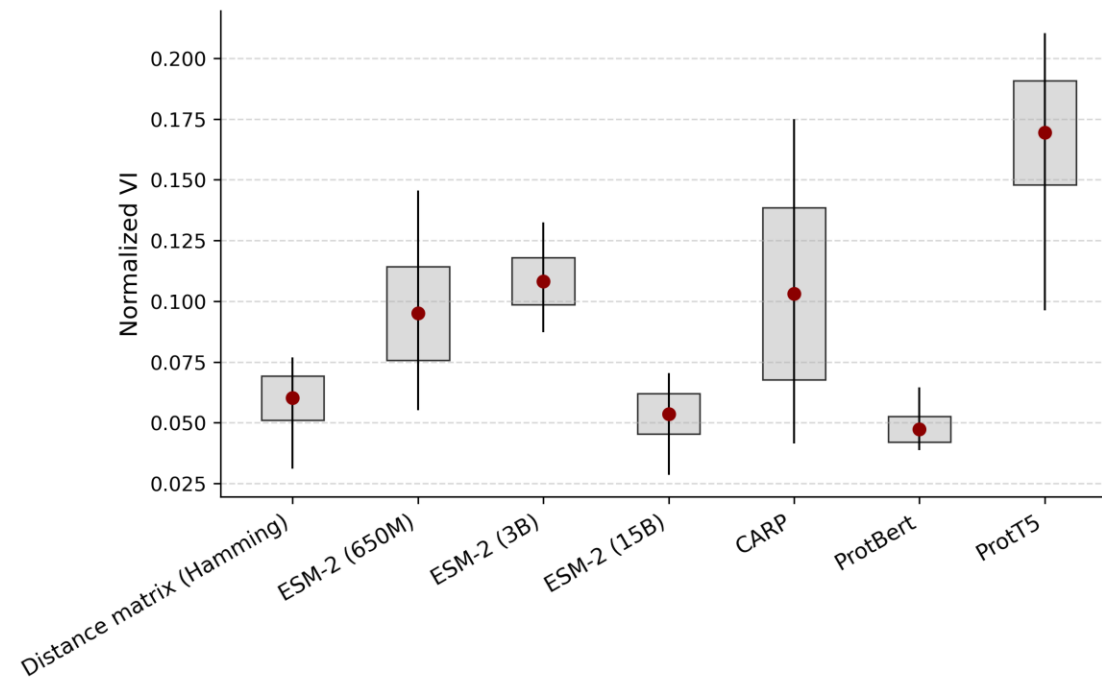
Parametras	Reikšmės
Įterpinio modelis	ProtT5, ESM2-[t33,t36,t48], ProtBert, ProtT5, CARP
Dimensijos mažinimo metodas	PCA, t-SNE, UMAP, MDS
Naudojama atstumu metrika	Euklidinis, kosinusų panašumas
Sumažintos dimensijos dydis	1,2,4,6,8,10

Rezultatai

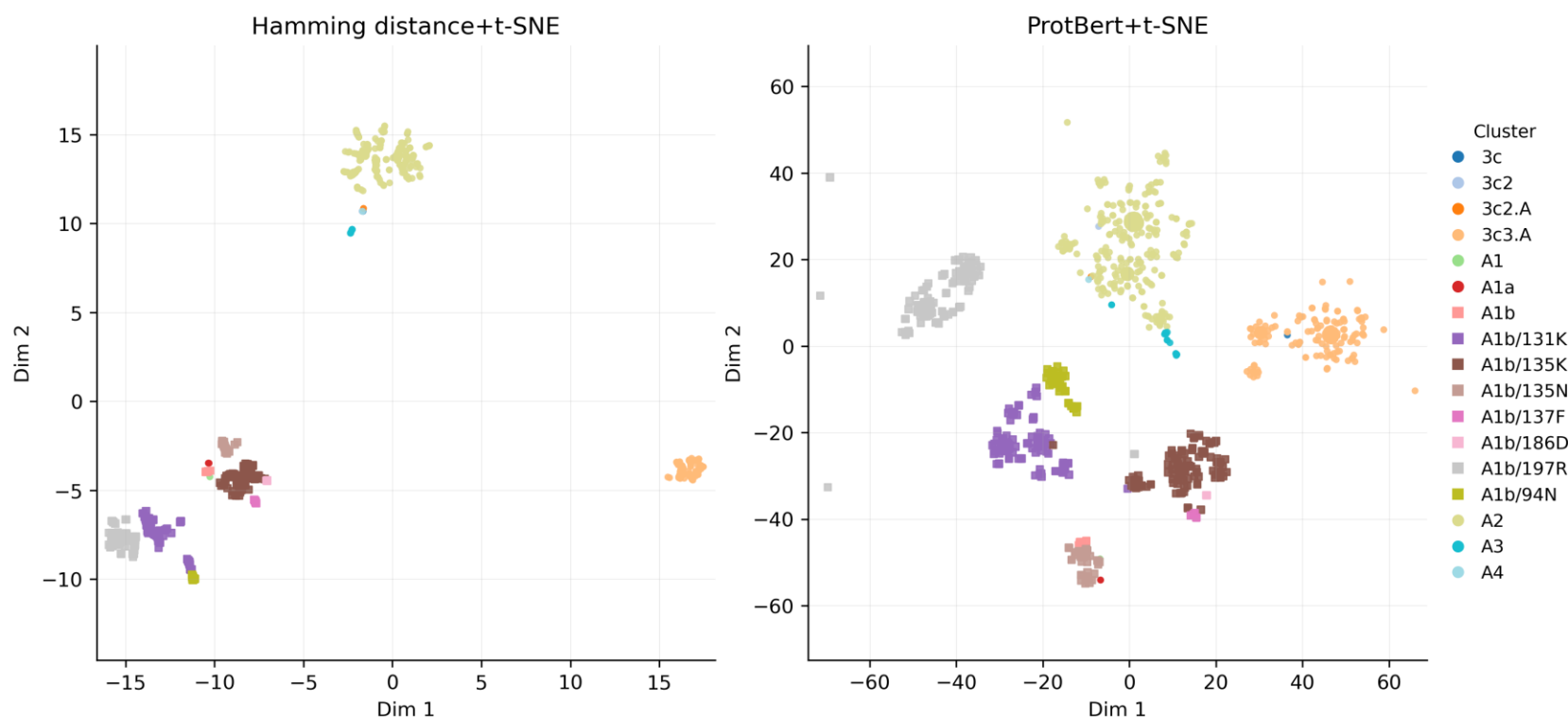
Metodas	Normalizuotas VI apmokyme	Normalizuotas VI validacijoje
t-SNE	0.0662	0.0530
UMAP	0.0815	0.0864
PCA	0.0944	0.0757
MDS	0.1052	0.0780
ESM-2[650M] + t-SNE	0.1214	0.0936
ESM-2[3B]+ t-SNE	0.1180	0.1106
ESM-2[15B]+ t-SNE	0.1095	0.0608
ProtT5 + t-SNE	0.1641	0.1384
ProtBert + t-SNE	0.1259	0.0492
CARP	0.1125	0.1036

Rezultatų stabilumas

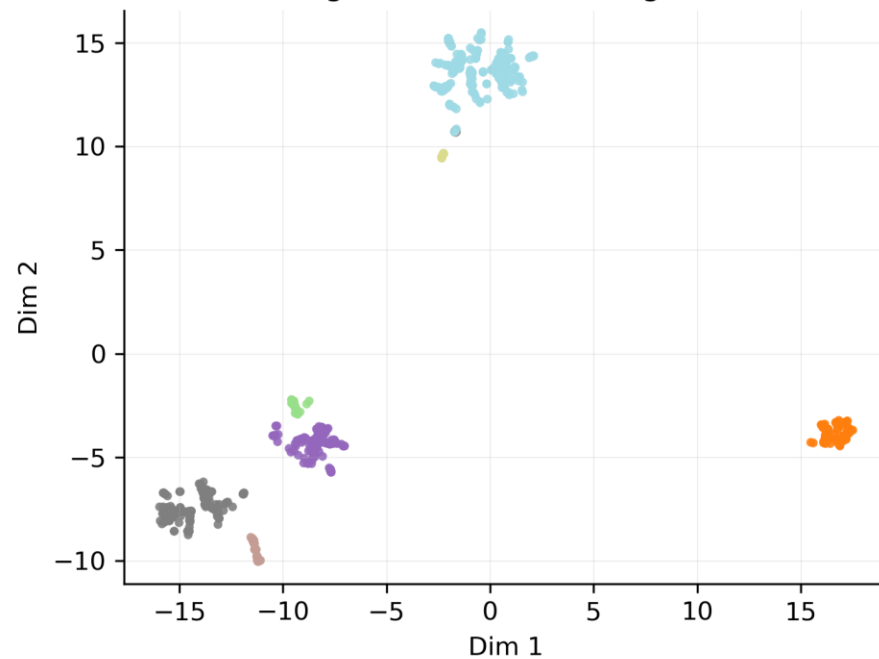
100 bandymų
eliminavus 10 procentų
atsitiktinių
duomenų sekų.



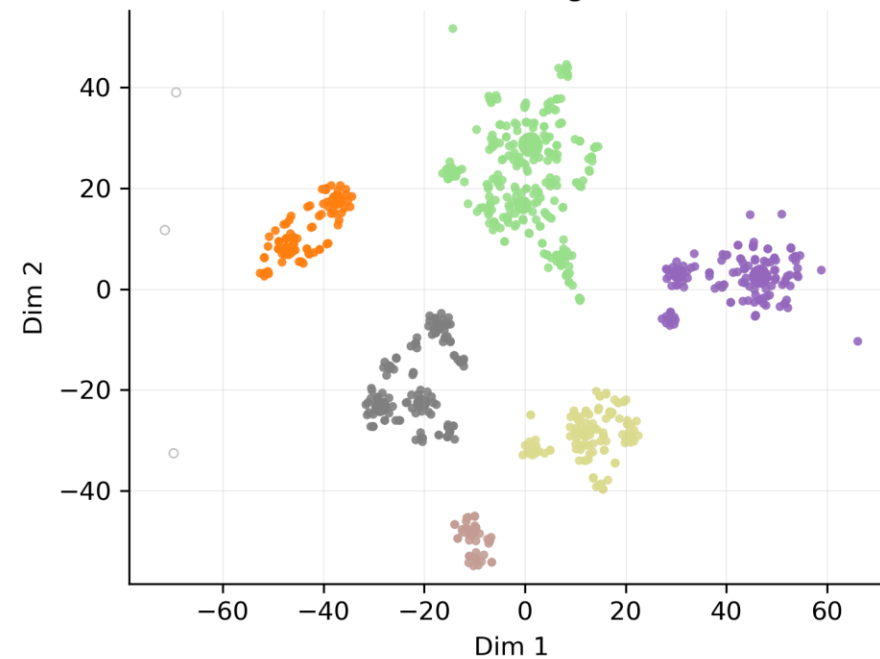
Įterpinių palyginimas



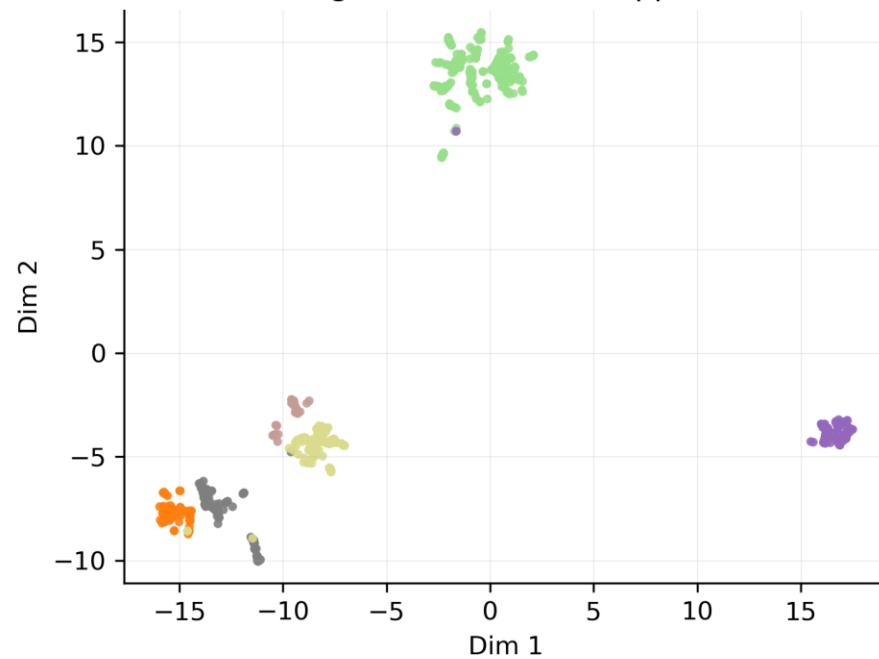
Hamming distance+t-SNE original labels



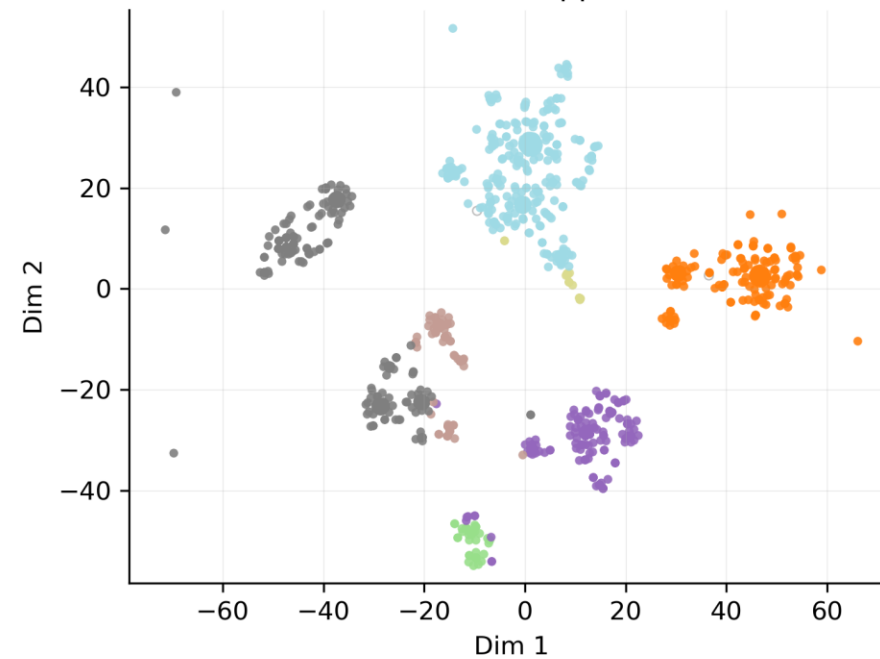
ProtBert+t-SNE original labels



Hamming distance+t-SNE flipped labels



ProtBert+t-SNE flipped labels



Kito pusmečio darbo planas

Mokslinė veikla:

- Gautų duomenų analizė, apibendrinimas, išvadų parengimas:
 - Teorinės tyrimo dalies apibendrinimas.
 - Empirinio tyrimo dalies apibendrinimas.
 - Rezultatų apibendrinimas, esminių rezultatų išskyrimas bei išvadų parengimas.

Mokslinė stažuotė

Mokslinė stažuotė Varšuvos universitete 10 01 – 12 31

Publikacijų rengimas

Publikacijos rengimas Computational Biology, BMC Bioinformatics ar panašiam leidinyje turinčiame Impact Factor.