

Niniejsza rozprawa stanowi część badania finansowanego przez Narodowe Centrum Badań i Rozwoju w Polsce (grant nr DOB-BIO10/03/01/2019) i przedstawia nowatorskie podejście do kryminalistycznej analizy gleby poprzez zastosowanie BioDataScience w metagenomice. Praca kładzie nacisk na integrację metod głębokiego uczenia oraz metod opartych na teorii informacji w celu rozwoju metod opartych na mikrobiologii środowiska w kryminalistyce. Kluczową rolę w formułowaniu założeń projektu odegrały integracja i adnotacja danych za pomocą metadanych, a także obszerne badania wstępne oraz eksploracyjna analiza danych dotyczących właściwości środowiskowych potencjalnych miejsc poboru próbek.

Głównym elementem tego badania jest zastosowanie metod głębokiego uczenia do analizy danych metagenomowych bez polegania na referencyjnych bazach danych taksonomicznych czy funkcyjnych. Poprzez konstrukcję reprezentacji w postaci grafów de Bruijna za pomocą narzędzia MetaGraph oraz kwantyfikację zaprojektowanych cech mikrobiologicznych, zastosowane podejście badawcze skutecznie wyodrębnia i identyfikuje charakterystyczne i informatywne markery mikrobiomu gleby, wykorzystując metadane w formie adnotacji. Opracowano nowatorski system głębokiego uczenia, wykorzystujący blisko dziesięcioletnią funkcję kosztu *Triplet Loss*, w celu rozróżniania subtelnych podobieństw i różnic między próbkami gleby. Podkreśla to, w jaki sposób ukryte właściwości danych mogą wspomagać bioinformatykę i pozwalać na zmniejszenie zależności od tradycyjnych baz danych. Zdolność tego systemu do transformacji cech i przypisywania prawdopodobieństw dotyczących geograficznego pochodzenia próbek skutecznie zwiększa predykcyjną dokładność analizy kryminalistycznej. Autorskie podejście zostało zwalidowane w porównaniu ze “złotym standardem” metod uczenia maszynowego na dobrze zdefiniowanych i czystych danych, a wyniki wykazały różnice między tymi metodami, dostosowanymi do danych, a systemem głębokiego uczenia skrojonym na miarę tych danych. Użyteczność systemu została również potwierdzona na próbkach gleby pochodzących z symulowanych spraw kryminalistycznych.

Integracja wyników tych badań z platformą SMAFT wyznacza nowy standard w kryminalistycznej analizie gleby i ilustruje praktyczne zastosowania zaawansowanych technik *BioData science* w warunkach kryminalistycznych. Niniejsza rozprawa podkreśla nieprzemijający potencjał wykorzystania teorii informacji i głębokiego uczenia w naukach o życiu.