Knowledge-embedded AI to improve the food quality and sustainability of food supply chains

Dr. Xuezhen Guo

Senior researcher, Data Science, Informatics & Supply Chains

Wageningen Food & Biobased Research

Jan 21, 2025

WAGENINGEN
UNIVERSITY & RESEARCH

Machine learning (AI)

Predictors

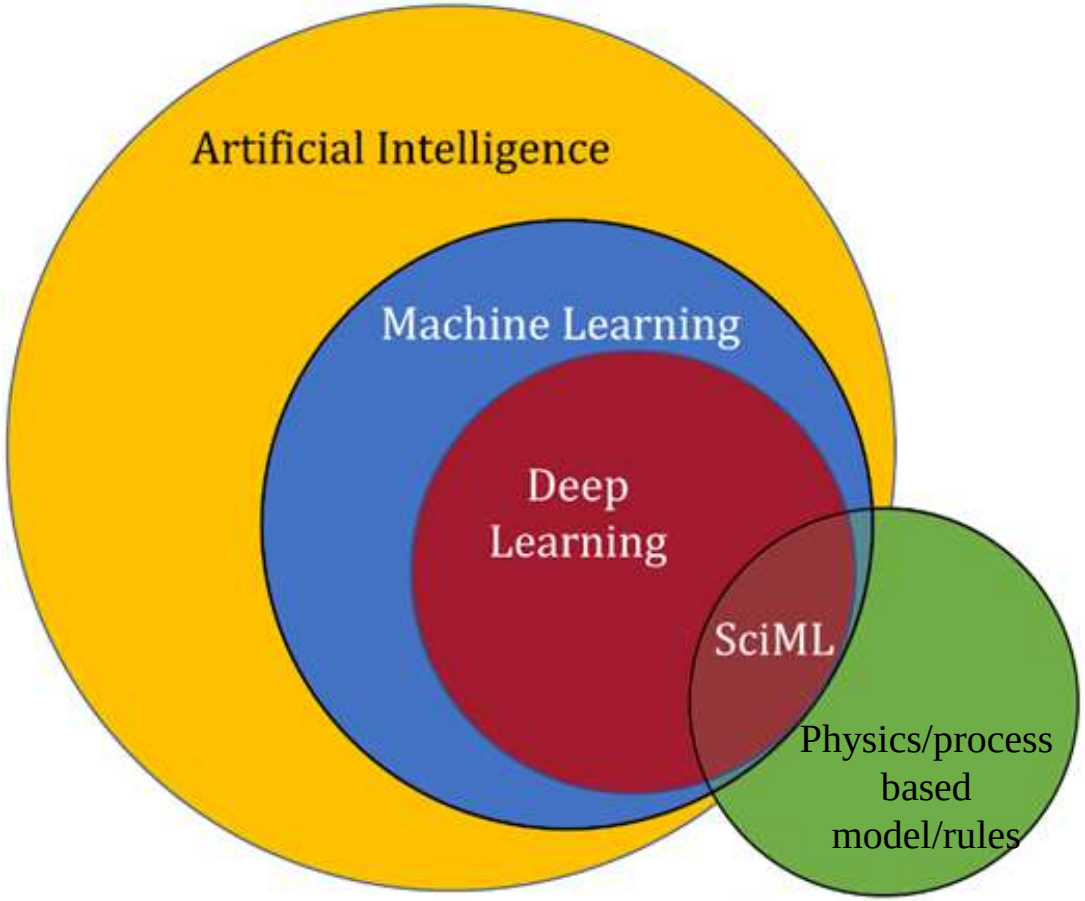
Data

ML model

Predicted targets

Scientific Machine Learning (SciML): knowledge-embedded AI

- Physics/process-based model/ rules + neural networks (deep learning).
- Lack of data
- Domain knowledge reduces data requirement for model training



https://sciml.wur.nl/reviews/sciml/sciml.html

Nobel prize 2024: the year of AI and SciML

The Nobel Prize in Physics 2024

Ill. Niklas Elmehed © Nobel Prize Outreach
John J. Hopfield
Prize share: 1/2

Ill. Niklas Elmehed © Nobel Prize Outreach
Geoffrey E. Hinton
Prize share: 1/2

The Nobel Prize in Chemistry 2024

Ill. Niklas Elmehed © Nobel Prize Outreach
David Baker
Prize share: 1/2

Ill. Niklas Elmehed © Nobel Prize Outreach
Demis Hassabis
Prize share: 1/4

Ill. Niklas Elmehed © Nobel Prize Outreach
John M. Jumper
Prize share: 1/4

Neural networks

Domain-knowledge based Computational model

SciML: Alphafold2

Differences between black box and SciML models

OpenAI (ChatGPT)

Alphafold2

What are the differences between these two models?

Parameters: hundreds of billions
Training samples: trillions of words
Type: Artificial General Intelligence (AGI)
Model complexity: **relatively low**

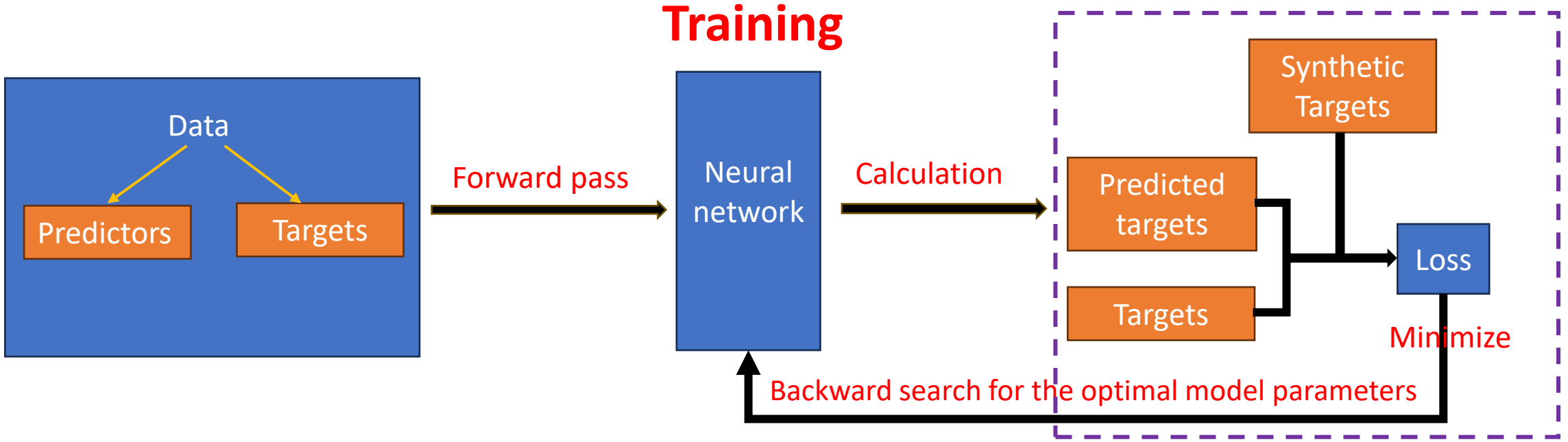
Parameters: millions
Training samples: < 120,000 protein structures
Type: Domain-Specific AI
Model complexity: **relatively high**

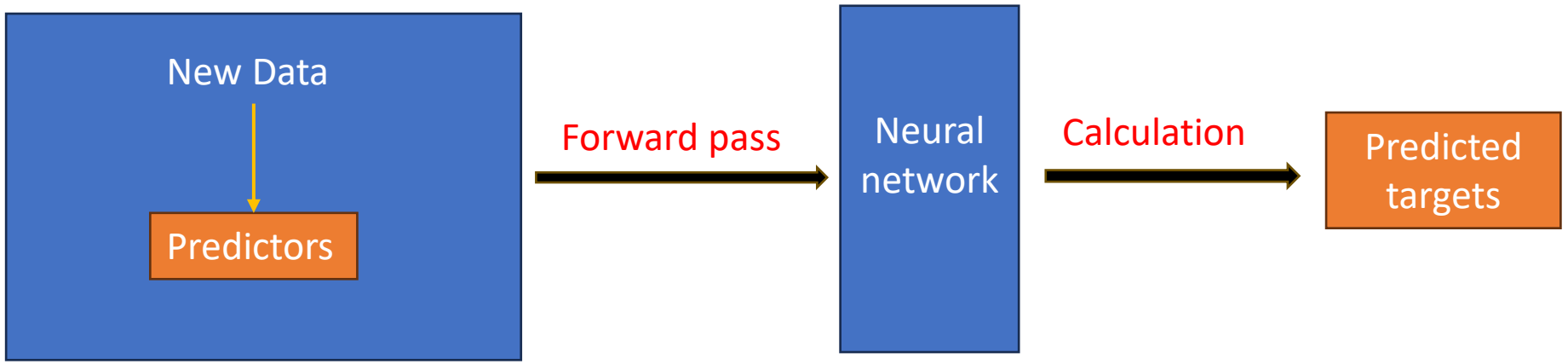
SciML model typology

- Physics-guided model
- Physics-encoded model
- Physics-informed model

Faroughi, S.A., et. al. (2024)

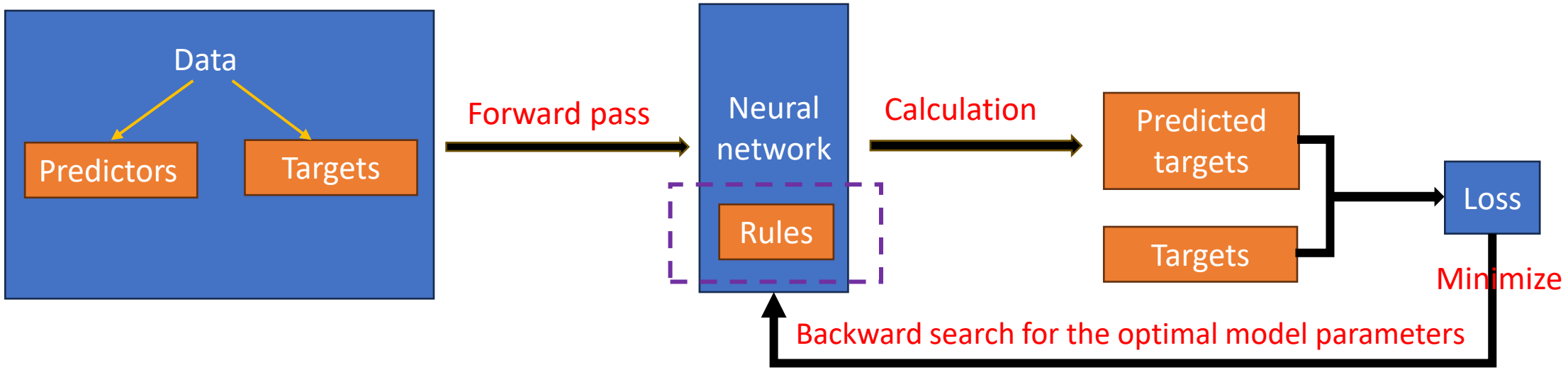
SciML: 1. Knowledge-guided model



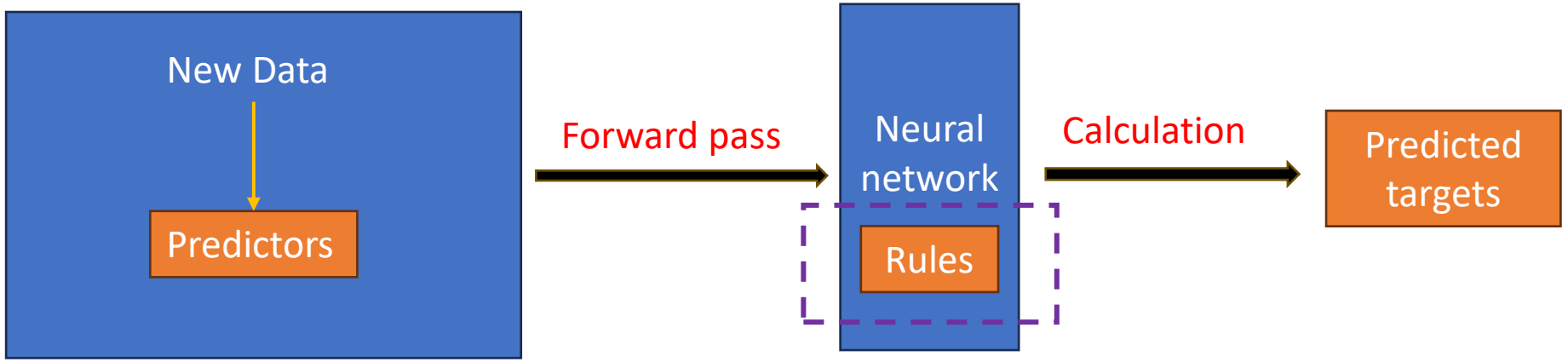


SciML: 2. Knowledge-encoded model

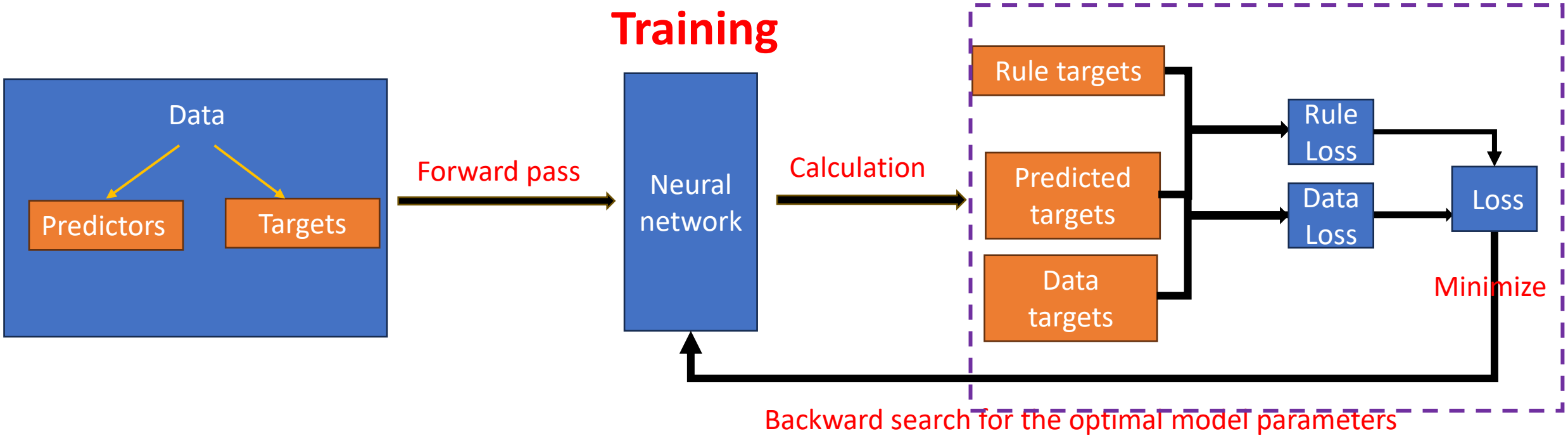
Training

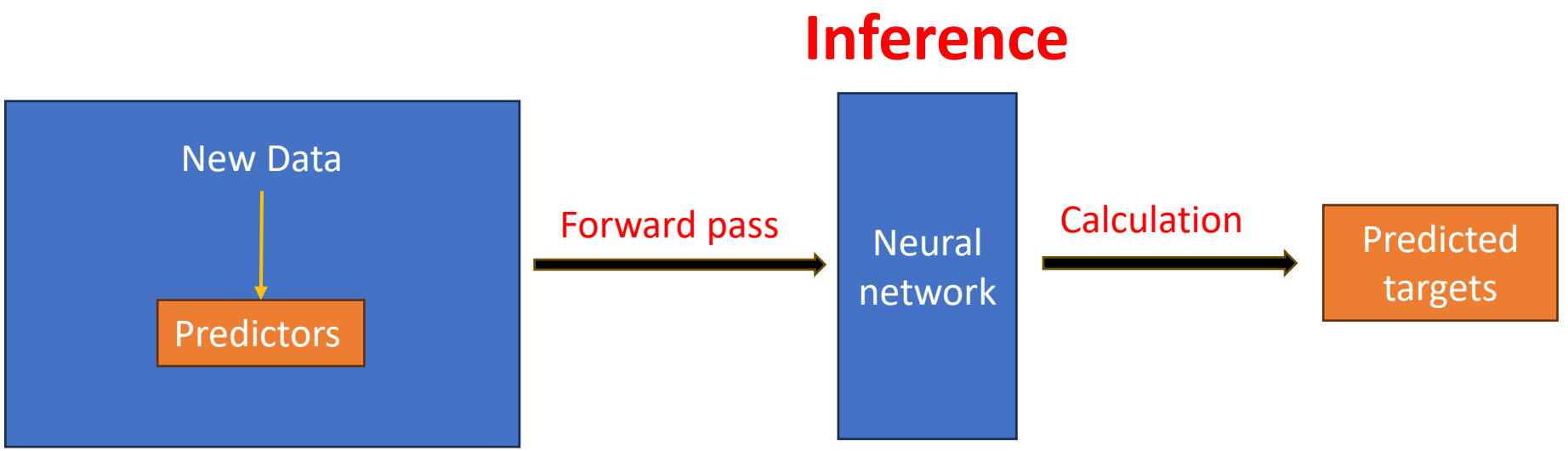


Inference



SciML: 3. Knowledge-informed model

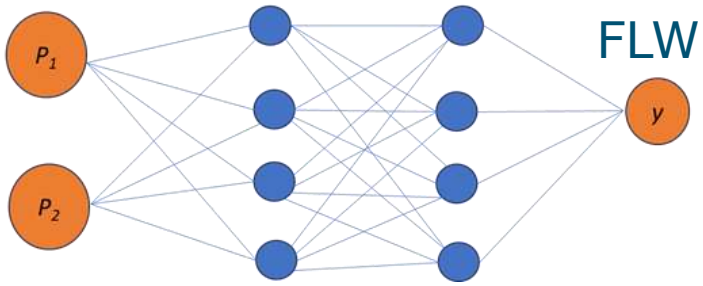




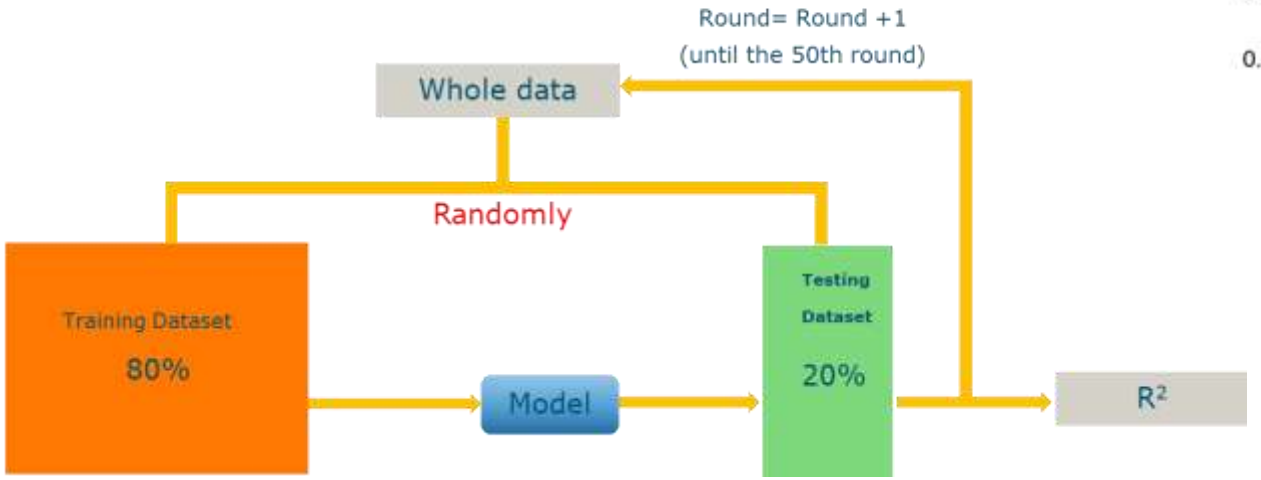
Food loss & waste (FLW) prediction in the food chains

Predictors = [“FAO Region”, “year”, “Population (1000 persons)” “GDP_pc”, “precipitation (mm)”, “Protein_supply (g/capita/day)”, “Production (1000t)”]

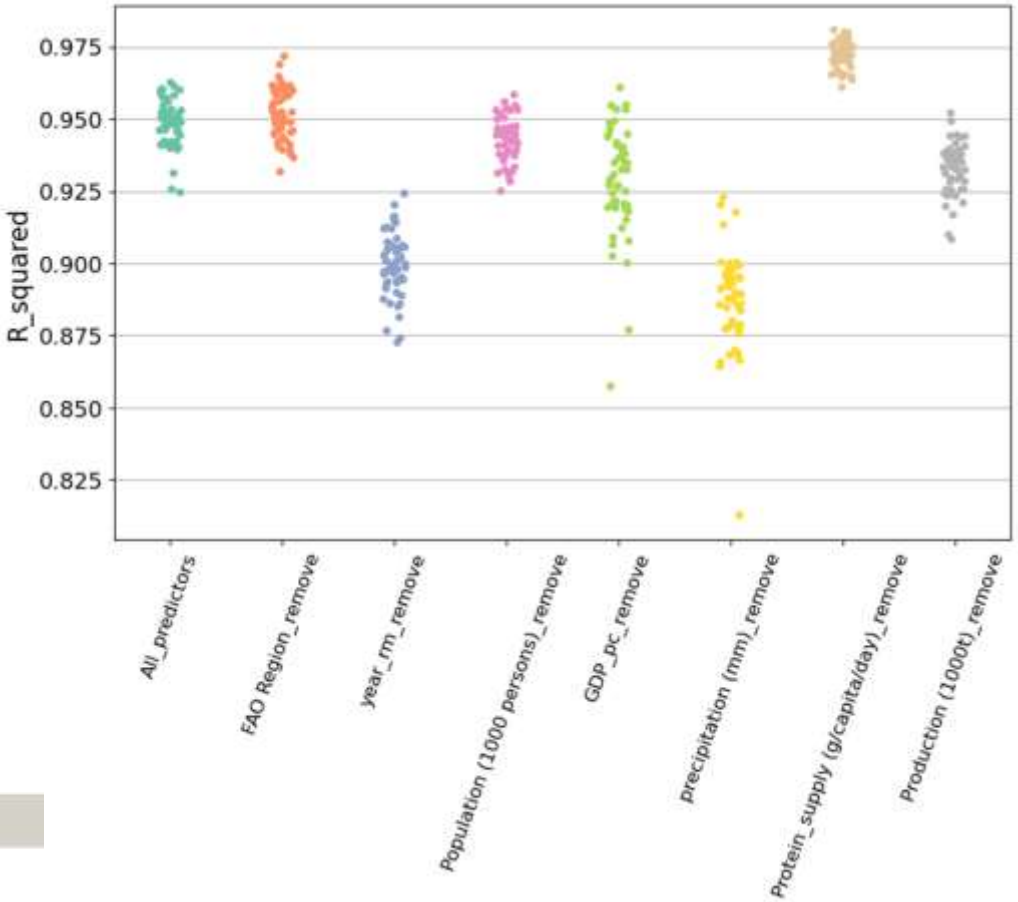
Simplified graphical representation



Training and testing:

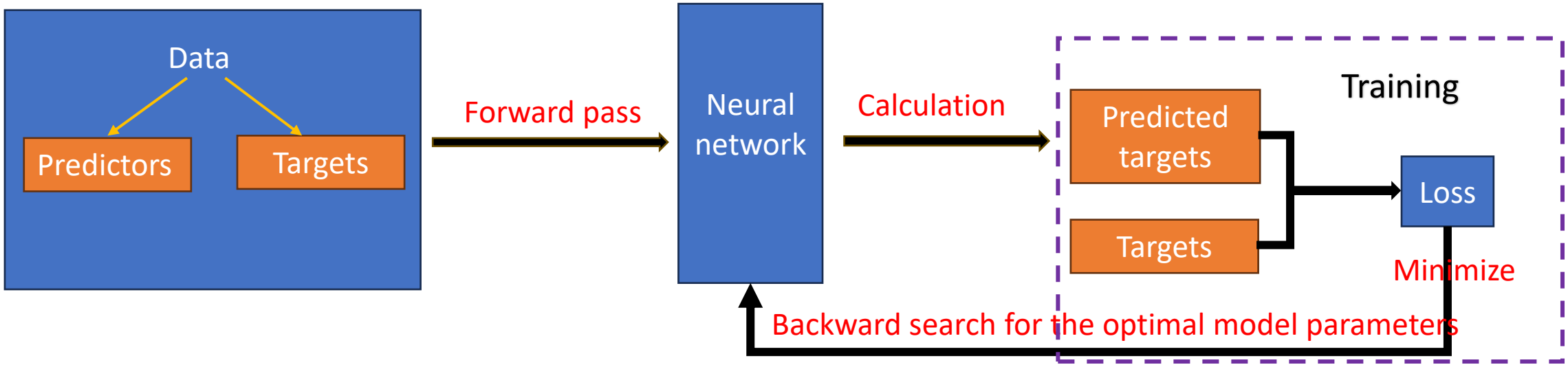


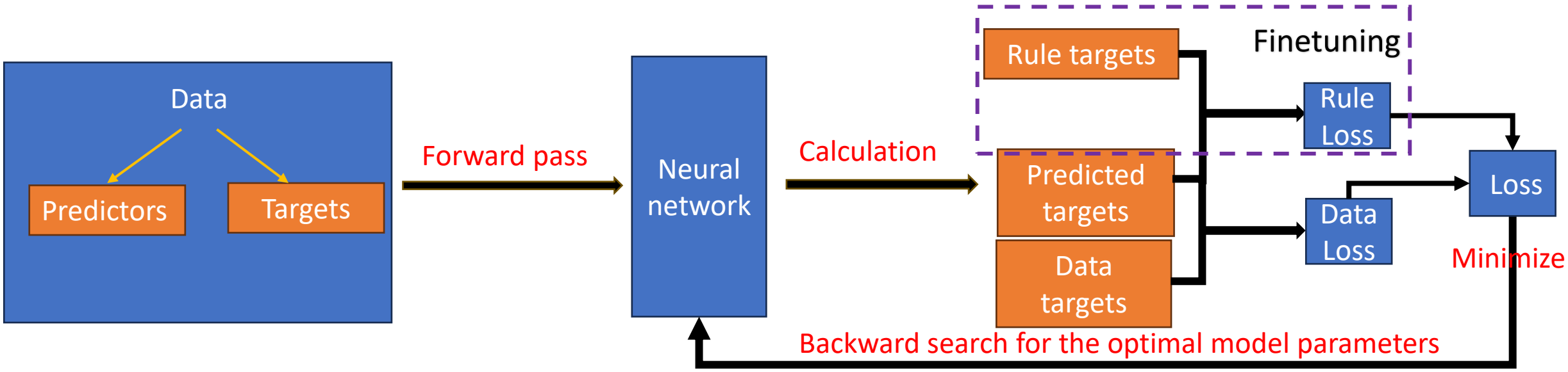
The prediction accuracy between actual and predicted FLW on the testing data using different sets of predictors, each for 50 rounds



WAGENINGEN
UNIVERSITY & RESEARCH

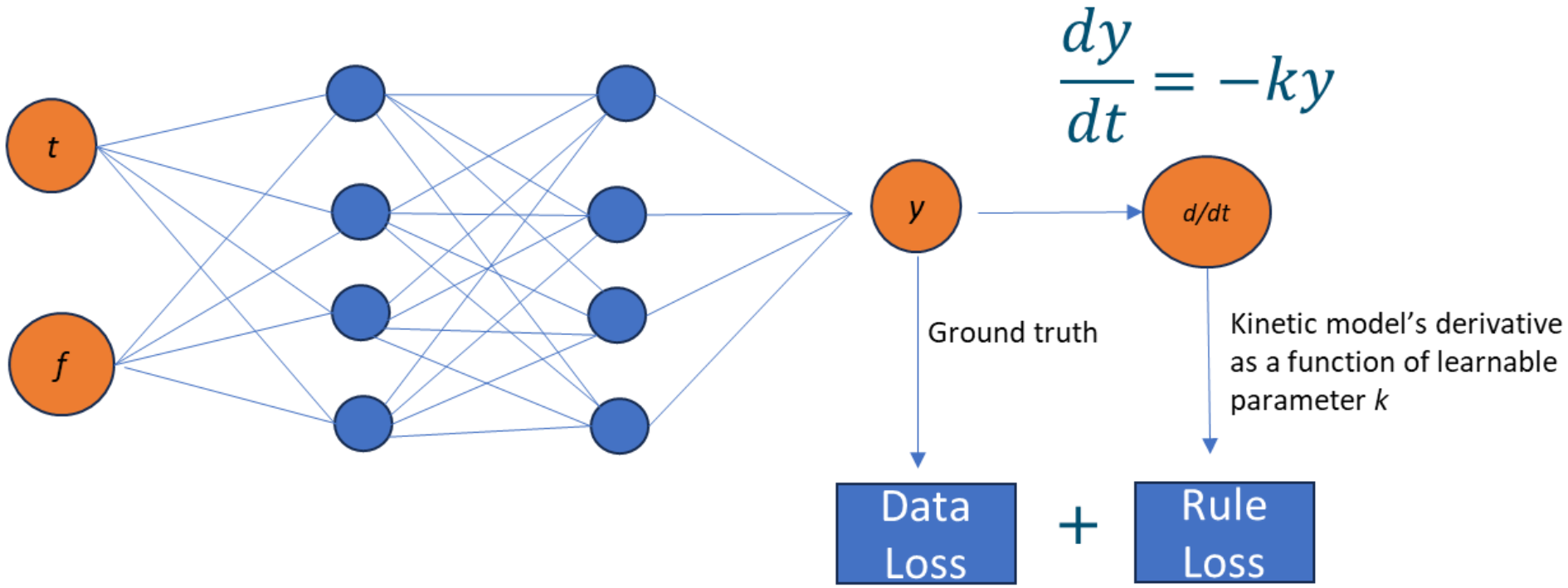
Scientific Machine Learning (SciML) to finetune the original model for FLW forecasting with knowledge-based rules (e.g. experts or other models)





SciML for quality prediction in the food chains

Simplified representation of the KiNN



SciML for taste prediction

Real target variable data

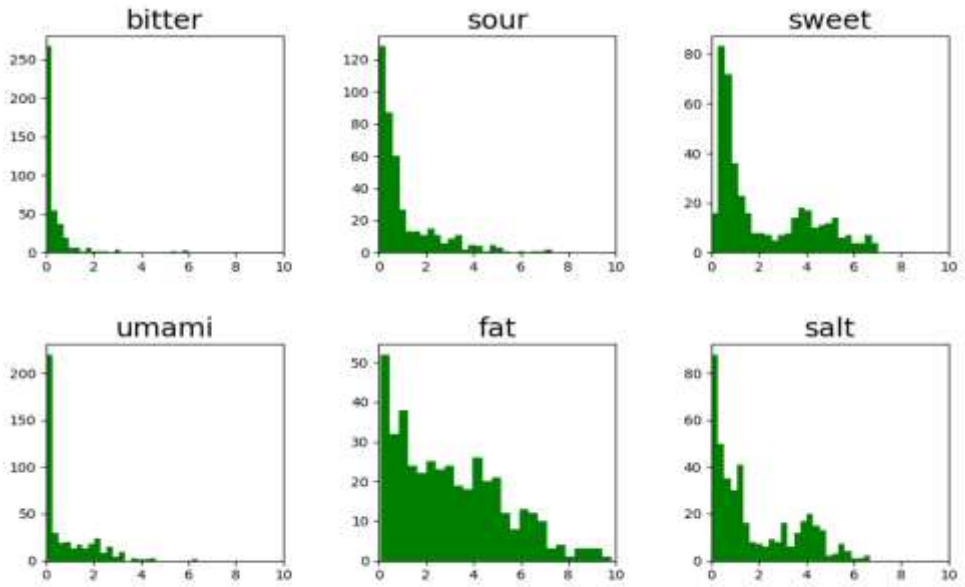


Figure 1. A histogram of the target variables of both the Sensory and INRAE data combined.

+

Synthetically generated target variable data with knowledge-based methods

A combined training dataset

Take away

SciML can help researchers in the food domain to employ AI to improve the efficiency and quality of their research, even when data availability is relatively limited.

Email contact:
xuezhen.guo@wur.nl