

GENE EXPRESSION

Gene expression is essential in gene prediction because it provides valuable insights into which genes are actively functioning and producing proteins within an organism's cells.

TRANSFER LEARNING

cross-species transfer learning is a machine learning technique that involves transferring knowledge learned from one species to another

Humans transfer learning across related tasks..

Domain: Driving



Task A: Driving a Bicycle

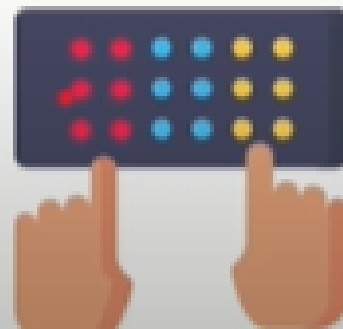


Task B: Driving a Scooter

Domain: Communication



Task A: Writing by hand



Task B: Typing on keyboard

Then

SNP and indel

large dataset with known gene expression levels and associated features is necessary.

However, obtaining a substantial amount of labeled data for a specific task in a particular species can be challenging and is often limited to species where such data is readily available

Now

RNN and CNN

Addresses the need for a large amount of labeled data by leveraging transfer learning and cross-species genomic information.

Improves gene expression predictions even in species with limited data, enhancing its versatility and applicability across a broader range of plant species.

DNA

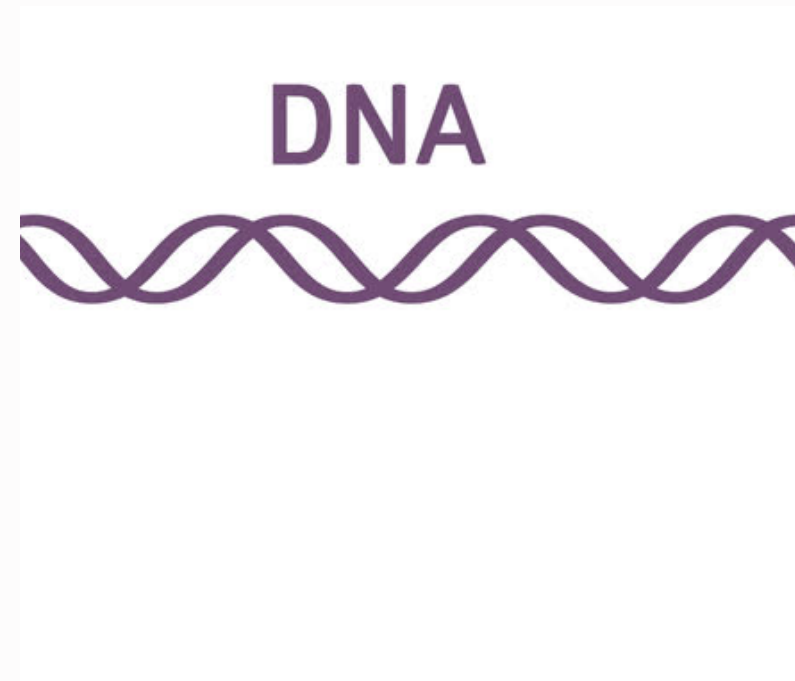
It is a double-stranded molecule, forming a double helix.

DNA contains the genetic information needed for gene prediction.

RNA

RNA is usually single-stranded, but it can fold into secondary structures by pairing its own bases within the same molecule

plays a crucial role in gene expression and the regulation of genes.



mRNA

It is a type of RNA that is also single-stranded, and it contains ribose sugar and the same bases as RNA

carries the genetic code from DNA to the ribosomes, where proteins are synthesized.

CORNBERT

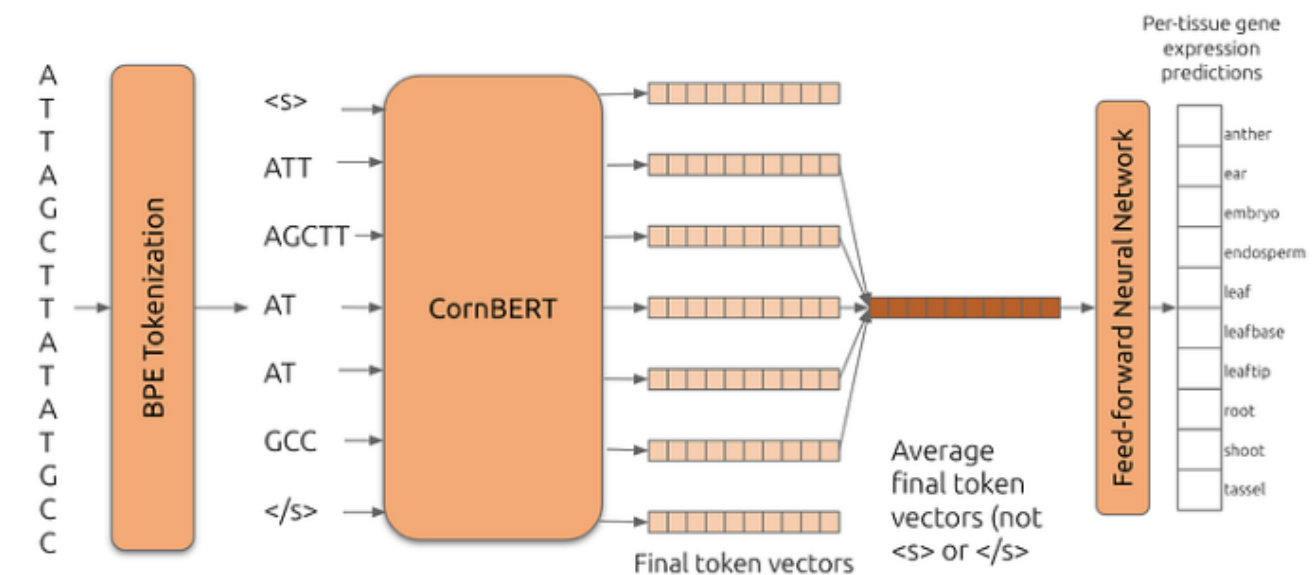
Corn Bert is a model very similar to Flora-bert which is very specific to Maize prediction.

They are using 3 databases MaizeGDB, RefSeq and Ensemble

Research paper:

<https://towardsdatascience.com/bringing-bert-to-the-field-how-to-predict-gene-expression-from-corn-dna-9287af91fcf8>

Github <https://github.com/ncgr/CornBER>



Schematic of pre-trained CornBERT with prediction head, including the byte-pair encoding tokenization step. Image by the authors.