PARKINSON'S DISEASE PROGRESSION PREDICTION

Ran Minerbi

INPUT DATA

- Raw medical data from [Kaggle]
  - Patient_id , visit_mounth , raw medical data

Train clinical data

```
55_0,55,0,10,6,15,,
55_3,55,3,10,7,25,,
55_6,55,6,8,10,34,,
55_9,55,9,8,9,30,0,On
55_12,55,12,10,10,41,0,On
55_18,55,18,7,13,38,0,On
55_24,55,24,16,9,49,0,On
55_30,55,30,14,13,49,0,On
55_36,55,36,17,18,51,0,On
55_42,55,42,12,20,41,0,On
55_48,55,48,17,16,52,0,On
55_54,55,54,12,18,51,0,On
55_60,55,60,23,21,56,0,On
942_0,942,0,3,2,20,,
942_3,942,3,7,2,17,,
942_6,942,6,8,2,21,,
942_12,942,12,5,2,25,0,
942_18,942,18,6,5,18,0,
942_24,942,24,2,3,23,,
942_30,942,30,4,4,19,0,
942_36,942,36,2,4,19,0,
```

train peptides

```
55_0,0,55,O00391,NEQEQPLGQWHLS,11254.3
55_0,0,55,O00533,GNPEPTFSWTK,102060.0
55_0,0,55,O00533,IEIPSSVQQVPTIIK,174185.0
55_0,0,55,O00533,KPQSAVYSTGSNGILLC(UniMod_4)EAEGEPQPTIK,27278.9
55_0,0,55,O00533,SMEQNGPGLEYR,30838.7
55_0,0,55,O00533,TLKIENVSYQDKGNYR,23216.5
55_0,0,55,O00533,VIAVNEVGR,170878.0
55_0,0,55,O00533,VMTPAVYAPYDVK,148771.0
55_0,0,55,O00533,VNGSPVDNHPFAGDVVFPR,55202.1
55_0,0,55,O00584,ELDLNSVLLK,27229.3
55_0,0,55,O00584,HGTC(UniMod_4)AAQVDALNSQKK,12356.5
55_0,0,55,O14498,ALPGTPVASSQPR,41526.9
55_0,0,55,O14773,LFGGNFAHQASVAR,24884.4
55_0,0,55,O14773,LYQQHGAGLFDVTR,6353.65
55_0,0,55,O14791,VTEPISAESGEQVER,4202.71
55_0,0,55,O15240,AYQGVAAPFPK,107076.0
55_0,0,55,O15240,QQETAAAETETR,3095.35
55_0,0,55,O15240,THLGEALAPLSK,67603.7
55_0,0,55,O15394,ASGSPEPAISWFR,39688.8
55_0,0,55,O15394,NIINSDGGPYVC(UniMod_4)R,23209.4
55_0,0,55,O43505,TALASGGVLDASGDYR,333376.0
55_0,0,55,O60888,TQSSLVPALTDFVR,166850.0
55_0,0,55,O75144,ALMSPAGMLR,54083.9
55_0,0,55,O75144,GLYDVVSVLR,44662.4
55_0,0,55,O75326,SEGLLAC(UniMod_4)GTNAR,6380.35
55_0,0,55,O94919,ILEVVNQIQDEER,44104.1
55_0,0,55,O94919,QALNTDYLDSDYQR,23463.5
55_0,0,55,P00441,ADDLGKGGNEESTKTGNAGSR,39389.3
```

train proteins

```
visit_id,visit_month,patient
55_0,0,55,O00391,11254.3
55_0,0,55,O00533,732430.0
55_0,0,55,O00584,39585.8
55_0,0,55,O14498,41526.9
55_0,0,55,O14773,31238.0
55_0,0,55,O14791,4202.71
55_0,0,55,O15240,177775.0
55_0,0,55,O15394,62898.2
55_0,0,55,O43505,333376.0
55_0,0,55,O60888,166850.0
55_0,0,55,O75144,98746.3
55_0,0,55,O75326,6380.35
55_0,0,55,O94919,67567.6
55_0,0,55,P00441,64117.8
55_0,0,55,P00450,1181230.0
55_0,0,55,P00734,688909.0
55_0,0,55,P00736,109541.0
55_0,0,55,P00738,3956470.0
55_0,0,55,P00746,111619.0
```

PROTEIN INPUT

- Protein input file
  - 232,742 lines of data
  - 1113 visits of 248 patients
  - 227 proteins and corresponding values

```
visit_id,visit_month,patient_id,UniProt,NPX
55_0,0,55,O00391,11254.3
55_0,0,55,O00533,732430.0
55_0,0,55,O00584,39585.8
55_0,0,55,O14498,41526.9
55_0,0,55,O14773,31238.0
55_0,0,55,O14791,4202.71
55_0,0,55,O15240,177775.0
55_0,0,55,O15394,62898.2
55_0,0,55,O43505,333376.0
55_0,0,55,O60888,166850.0
55_0,0,55,O75144,98746.3
55_0,0,55,O75326,6380.35
55_0,0,55,O94919,67567.6
55_0,0,55,P00441,64117.8
55_0,0,55,P00450,1181230.0
55_0,0,55,P00734,688909.0
55_0,0,55,P00736,109541.0
55_0,0,55,P00738,3956470.0
55_0,0,55,P00746,111619.0
55_0,0,55,P00747,347865.0
55_0,0,55,P00748,71835.4
55_0,0,55,P00751,637630.0
55_0,0,55,P01008,2676370.0
55_0,0,55,P01009,14415900.0
55_0,0,55,P01011,2025890.0
55_0,0,55,P01019,1984650.0
55_0,0,55,P01023,1953020.0
55_0,0,55,P01024,3916980.0
55_0,0,55,P01031,13033.3
```

PEPTIDES DATA

- Peptide input file
  - 1113 visits of 248 patients
  - 227 different types of proteins and 968 different peptides within abundance per each

```
55_0,0,55,O00391,NEQEQPLGQWHLS,11254.3
55_0,0,55,O00533,GNPEPTFSWTK,102060.0
55_0,0,55,O00533,IEIPSSVQQVPTIIK,174185.0
55_0,0,55,O00533,KPQSAVYSTGSNGILLC(UniMod_4)EAEGEPQPTIK,27278.9
55_0,0,55,O00533,SMEQNGPGLEYR,30838.7
55_0,0,55,O00533,TLKIENVSYQDKGNYR,23216.5
55_0,0,55,O00533,VIAVNEVGR,170878.0
55_0,0,55,O00533,VMTPAVYAPYDVK,148771.0
55_0,0,55,O00533,VNGSPVDNHPFAGDVVFPR,55202.1
55_0,0,55,O00584,ELDLNSVLLK,27229.3
55_0,0,55,O00584,HGTC(UniMod_4)AAQVDALNSQKK,12356.5
55_0,0,55,O14498,ALPGTPVASSQPR,41526.9
55_0,0,55,O14773,LFGGNFAHQASVAR,24884.4
55_0,0,55,O14773,LYQQHGAGLFDVTR,6353.65
55_0,0,55,O14791,VTEPISAESGEQVER,4202.71
55_0,0,55,O15240,AYQGVAAPFPK,107076.0
55_0,0,55,O15240,QQETAAAETETR,3095.35
55_0,0,55,O15240,THLGEALAPLSK,67603.7
55_0,0,55,O15394,ASGSPEPAISWFR,39688.8
55_0,0,55,O15394,NIINSDGGPYVC(UniMod_4)R,23209.4
55_0,0,55,O43505,TALASGGVLDASGDYR,333376.0
55_0,0,55,O60888,TQSSLVPALTDFVR,166850.0
55_0,0,55,O75144,ALMSPAGMLR,54083.9
55_0,0,55,O75144,GLYDVVSVLR,44662.4
55_0,0,55,O75326,SEGLLAC(UniMod_4)GTNAR,6380.35
55_0,0,55,O94919,ILEVVNQIQDEER,44104.1
55_0,0,55,O94919,QALNTDYLDSDYQR,23463.5
55_0,0,55,P00441,ADDLGKGGNEESTKTGNAGSR,39389.3
55_0,0,55,P00441,TLVVHEKADDLGKGGNEESTK,24728.6
```

CLINICAL DATA

- Clinical data input file
  - 2615 visits of 248 patients
  - Updrs[1-4] – Unified Parkinson's Disease Rating Scale clinical assessment for
    - **Mentation, Behavior, and Mood -** cognitive impairments
    - **Activities of Daily Living (ADL) -** Evaluates the patient's ability to perform everyday tasks
    - **Motor -** comprehensive assessment of motor function
    - **Complications of Therapy -** complications arising from medication
    - Clinical state on medication

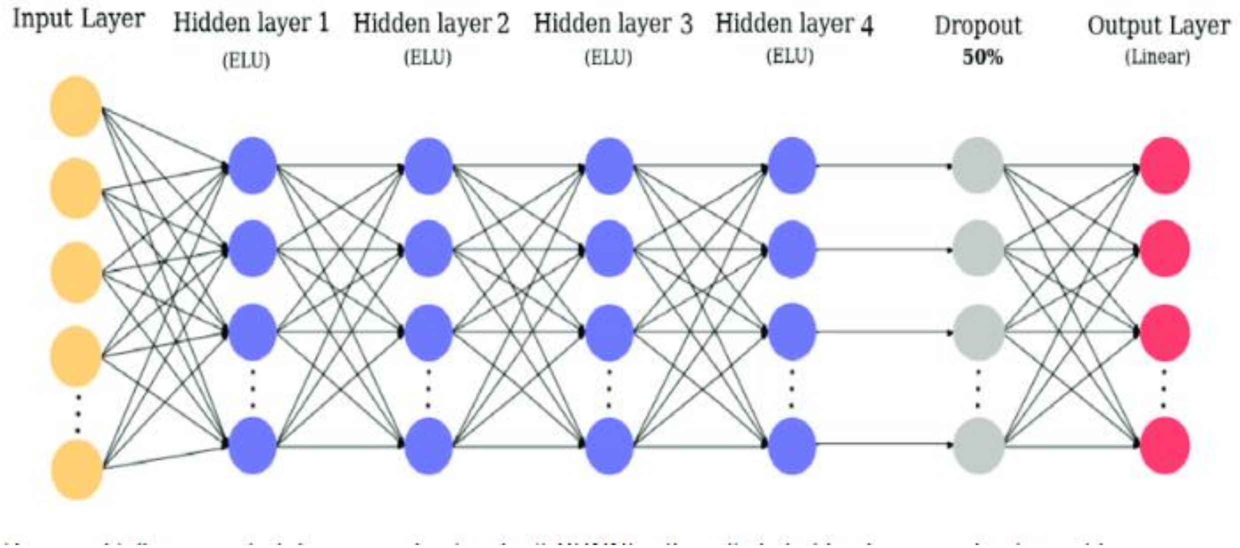
```
55_0,55,0,10,6,15,,
55_3,55,3,10,7,25,,
55_6,55,6,8,10,34,,
55_9,55,9,8,9,30,0,On
55_12,55,12,10,10,41,0,On
55_18,55,18,7,13,38,0,On
55_24,55,24,16,9,49,0,On
55_30,55,30,14,13,49,0,On
55_36,55,36,17,18,51,0,On
55_42,55,42,12,20,41,0,On
55_48,55,48,17,16,52,0,On
55_54,55,54,12,18,51,0,On
55_60,55,60,23,21,56,0,On
942_0,942,0,3,2,20,,
942_3,942,3,7,2,17,,
942_6,942,6,8,2,21,,
942_12,942,12,5,2,25,0,
942_18,942,18,6,5,18,0,
942_24,942,24,2,3,23,,
942_30,942,30,4,4,19,0,
942_36,942,36,2,4,19,0,
```

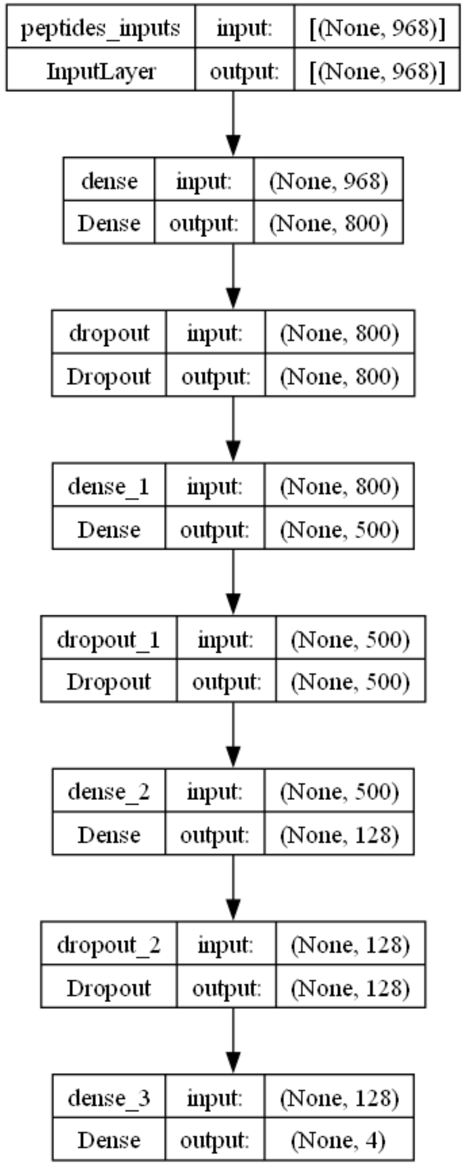
- # Our task
  - Predict updrs[1-4] clinical assesment after 4 years from first diagnose (mounth 48) based on patient medial history raw data

- # Data constraints
  - Many missing parameters or updrs clinical assessments
  - we have blood samples (protein/peptide) data for ~50% of visits
  - Updrs assessments provided by neurologists – may not be precise values

APPROACH 1

Predict updrs[1-4] for month 48 by peptide input data within **fully connected** neural network





TRAINING LOSS OUTLINE PROCEDURE

- Loss should be minimized during training

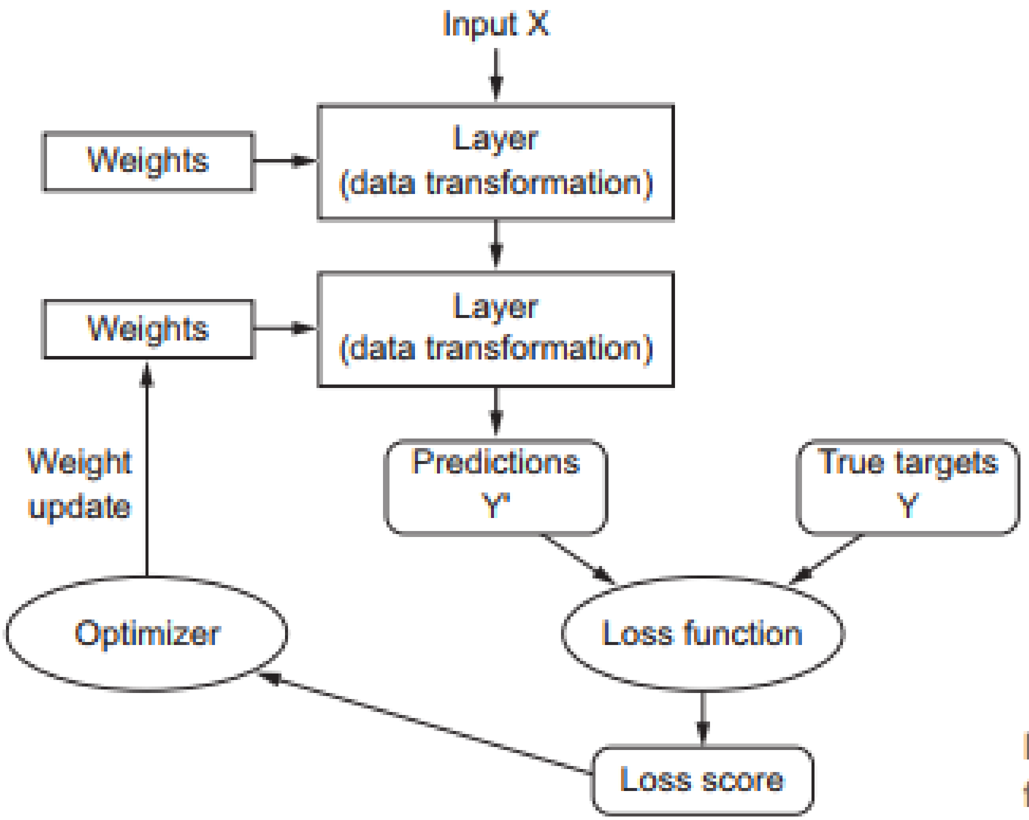
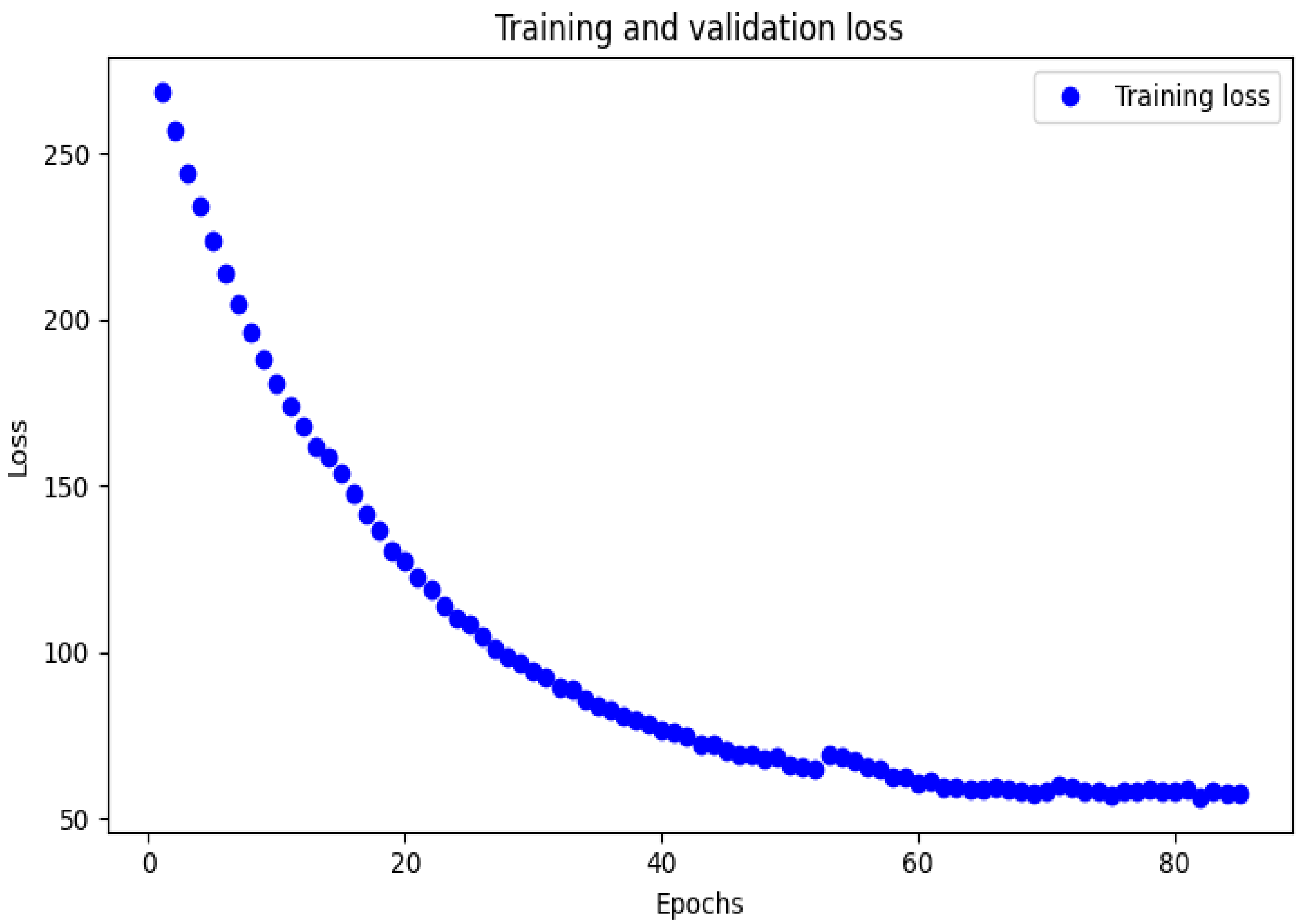


Figure 1.9 The loss score is used as a feedback signal to adjust the weights.

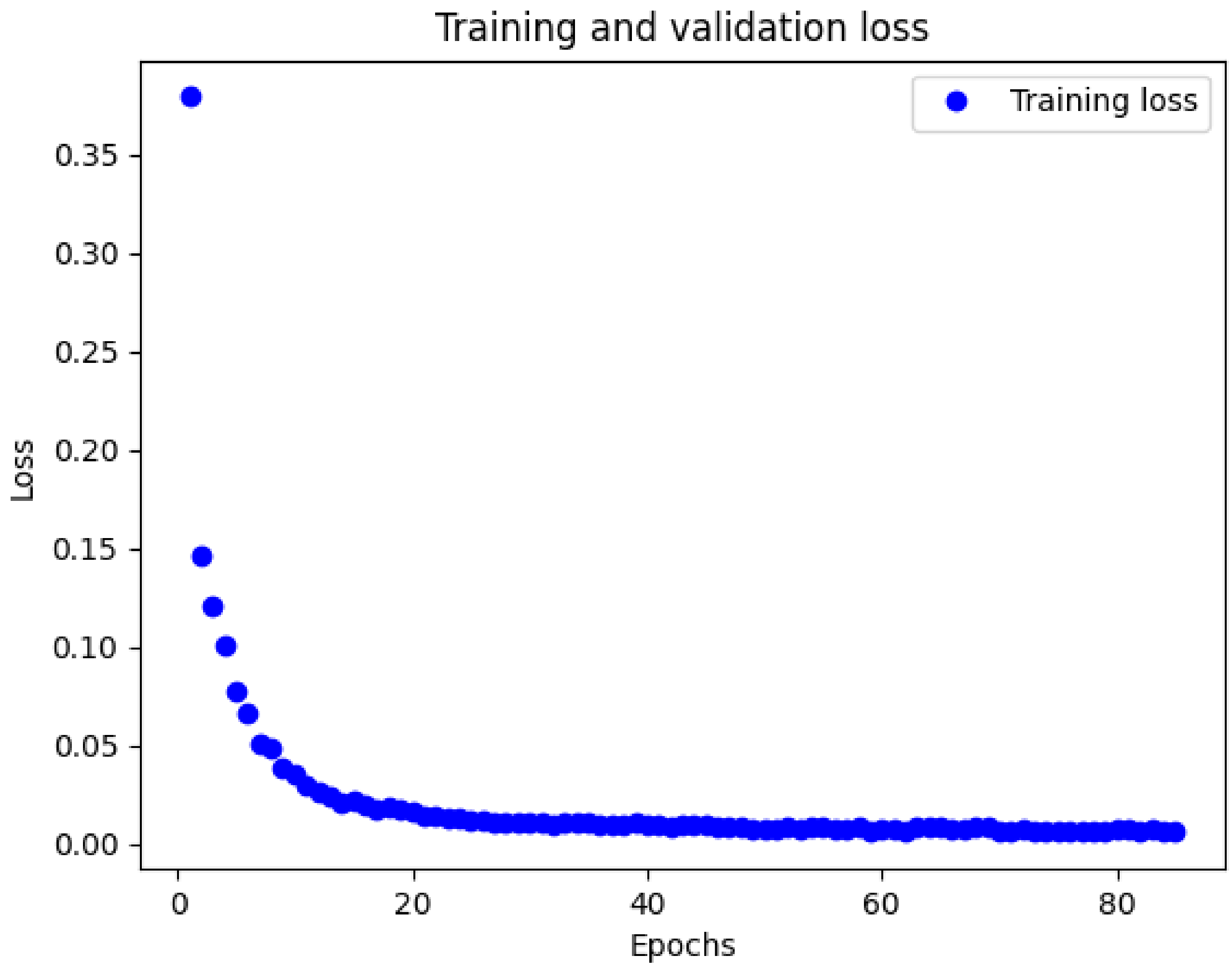
TRAINING LOSS

- Fit coverage failed - why ?



TRAINING LOSS WITHIN DATA NORMALIZATION

- Data is much more important than algorithm in use



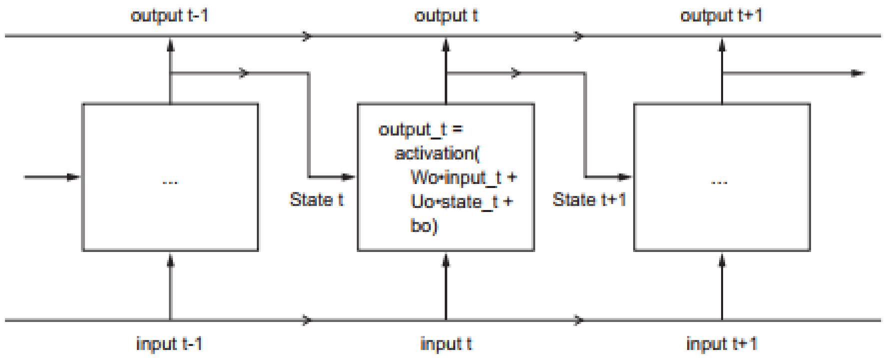
OVERFIT

- *Optimization* refers to the process of adjusting a model to get the best performance possible on the training data, whereas *generalization* refers to how well the trained model performs on data it has never seen before

- Learning how to deal with overfitting is essential to mastering machine learning

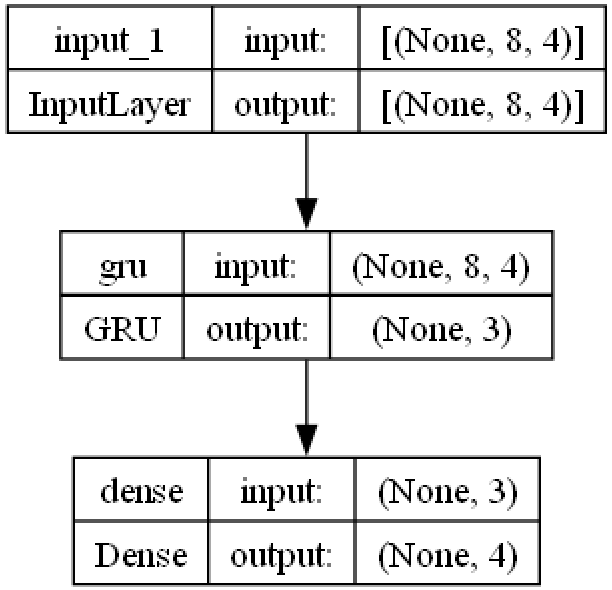
- **Avoid overfitting :**
  - Get more training data
  - Reduce network capacity
  - Add weight regularization
  - Add dropouts

LONG SHORT-TERM MEMORY (LSTM) NETWORK



APPROACH 2

Predict updrs[1-4] for month 48 by previous updrs assessments using **GRU** network



KERAS FUNCTIONAL API

- Merge layers originated by different architectures into single output prediction

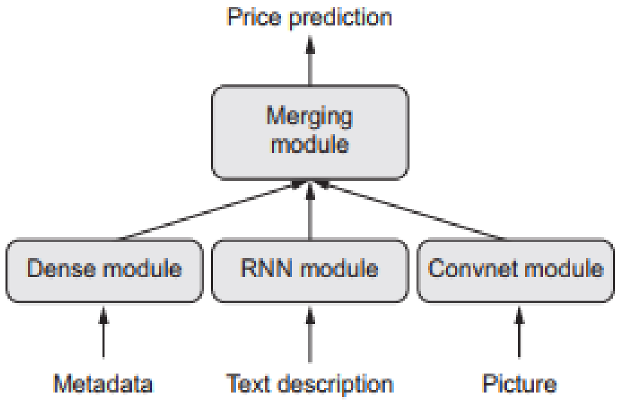


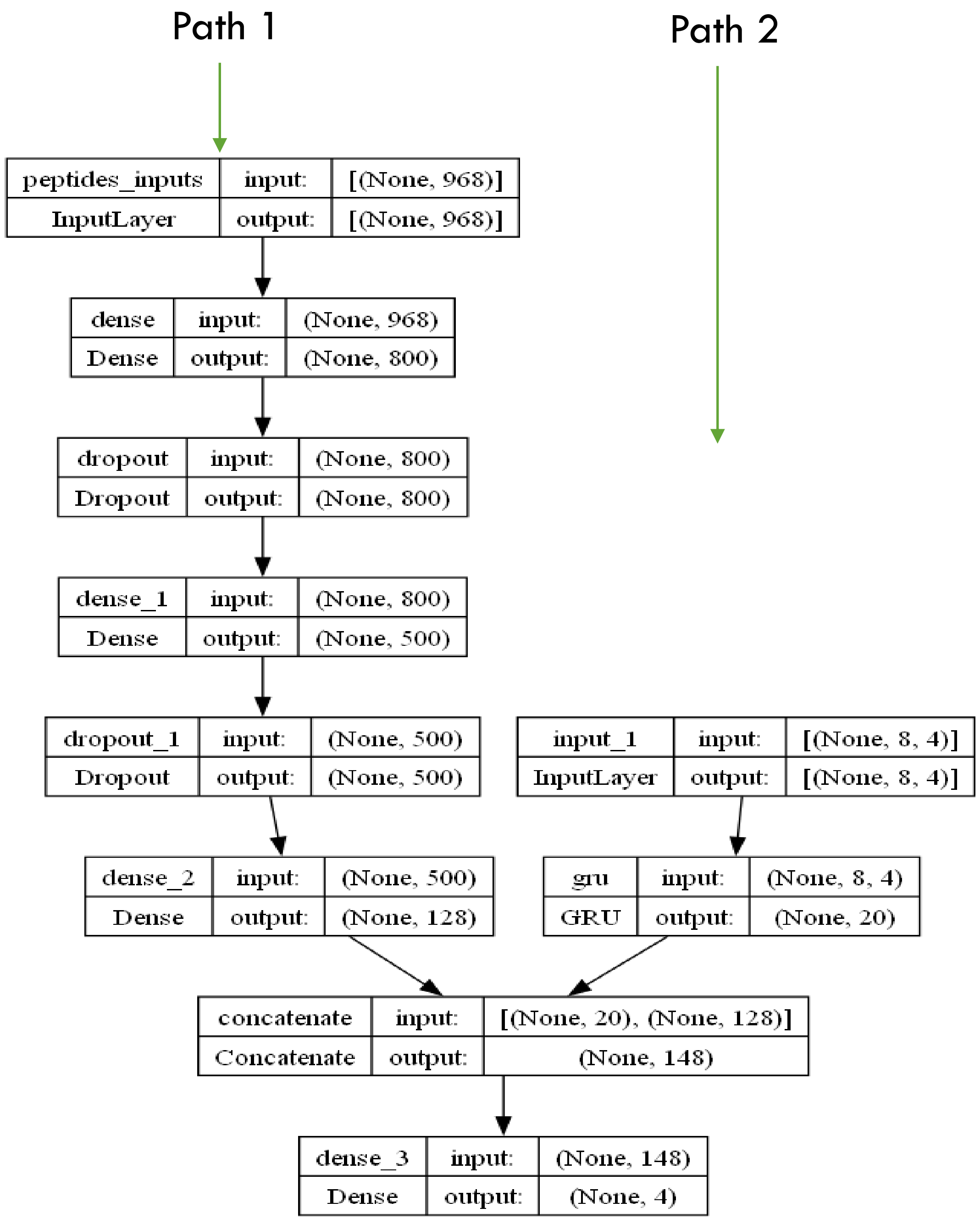
Figure 7.2 A multi-input model

APPROACH 3

Keras Functional API

- Path 1
  - Predict updrs for V_48 by peptide vector
- Path 2
  - Predict updrs for V_48 by previous updrs diagnostics (V_0, V_6, V_18, V_24, V36)

  - Merge layers and predict updrs for V_48



APPROACH 4

- Path 1
  - Predict updrs for V_48 by peptide vector
- Path 2
  - Predict updrs for V_48 by previous updrs diagnostics (V_0, V_6, V_18, V_24, V36)
- Path 3
  - Predict updrs for V_48 by protein blood samples from (V_0, V_18, V_36)

  - Merge layers and predict updrs for V_48