# NTxPred2: Detailed Instructions for Neurotoxicity Prediction

This document provides detailed instructions on using the NTxPred2 software for predicting the neurotoxicity of peptides and proteins. It covers installation, usage, and replication of results.

## 📌 Introduction

**NTxPred2** is a computational tool designed to predict the neurotoxic potential of peptides and proteins. It is particularly useful in the field of therapeutic peptide and protein development. The method employs large language model word embeddings as features for predicting neurotoxic activity. NTxPred2 offers Prediction, Protein-Scanning, and Design modules.

🔗 **NTxPred2 Web Server:** <http://webs.iiitd.edu.in/raghava/ntxpred2>

📖 **Reference:** Rathore et al. *A large language model for predicting neurotoxic peptides and neurotoxins.* **#Coming Soon#**.

## 🧪 Reproducibility Guide

This guide will help you replicate the core results of the NTxPred2 paper.

### 1. Software and Dependencies

NTxPred2 is written in Python 3 and requires several libraries. It is highly recommended to use a Conda environment to manage these dependencies.

#### 1.1 Conda Installation (Recommended)

If you don't have Conda, download and install it from <https://conda.io/>.

#### 1.2 Create Conda Environment

Using the provided environment.yml file:

git clone https://github.com/raghavagps/ntxpred2.git  
cd ntxpred2  
conda env create -f environment.yml  
conda activate NTxPred2

This will create a Conda environment named "NTxPred2" with all the necessary dependencies.

#### 1.3 Manual Dependency Installation (Alternative)

If you choose not to use Conda, ensure you have Python 3.10.7 and install the following packages using pip:

pip install scikit-learn==1.5.2  
pip install pandas==1.5.3  
pip install numpy==1.25.2  
pip install torch==2.1.0  
pip install transformers==4.34.0  
pip install joblib==1.4.2  
pip install onnxruntime==1.15.1  
pip install Biopython==1.81  
pip install tqdm==4.64.1

### 2. Download Pre-trained Models

The pre-trained models are essential for NTxPred2 to function.

1. Go to the download page: <https://webs.iiitd.edu.in/raghava/ntxpred2/download.html>
2. Download the model ZIP file.
3. Extract the contents of the ZIP file into the root directory of the NTxPred2 installation (the directory where you cloned the GitHub repository). Ensure the model files are directly accessible by the python scripts.

### 3. Run NTxPred2

Now you can run NTxPred2 to make predictions.

#### 3.1 Basic Prediction

To run a basic prediction on the provided example FASTA file:

python ntxpred2.py -i example.fasta -o output.csv -m 1 -j 1 -wd .

* -i example.fasta: Specifies the input file.
* -o output.csv: Specifies the output file name.
* -m 1: Selects model 1 (ESM2-t30, the peptide model).
* -j 1: Selects job type 1 (Prediction).
* -wd .: Sets the current directory as the working directory.

#### 3.2 Understanding the Output

The results will be saved in output.csv. This file will contain the input sequence ID and the predicted neurotoxicity (e.g., probability of being neurotoxic).

## 🛠️ Installation Options

### 🔹 PIP Installation

NTxPred2 can be installed via PIP:

pip install ntxpred2  
ntxpred2 -h # To see available options

**Important:** You will still need to download the pre-trained models as described in Step 2 above, even if you install via PIP. The PIP installation only provides the Python scripts, not the model files.

### 🔹 Standalone Installation

This involves cloning the repository and setting up the environment manually, as described in the Reproducibility Guide.

## ⚠️ Important Notes

* **Model Files:** The pre-trained models are not included in the GitHub repository due to their size. You **must** download them separately from the provided link and extract them into the correct directory.
* **Working Directory:** It is crucial to set the working directory (-wd option) correctly. This is where NTxPred2 will save the output files.

## 🔬 Classification Models

NTxPred2 uses different models for different sequence lengths:

* **ESM2-t30 (Peptide Model):** For sequences 7-50 amino acids. (-m 1)
* **ET (Protein Model):** For sequences ≥ 51 amino acids. (-m 2)
* **ET (Combined Model):** For mixed-length sequences. (-m 3)
* **Default Model:** ESM2-t30 (Peptide Model).

## 🚀 Usage Details

### 🔹 Command-Line Options

Here's a comprehensive overview of the command-line options:

usage: ntxpred2.py [-h] [-i INPUT] [-o OUTPUT] [-t THRESHOLD]  
 [-j {1,2,3,4}] [-m {1,2,3}] [-d {1,2}] [-wd WORKING DIRECTORY]  
 [-p POSITION] [-r RESIDUES] [-w {8-20}]  
  
NTxPred2: Neurotoxicity Prediction Tool  
  
Optional arguments:  
 -h, --help show this help message and exit  
 -i INPUT Input: Peptide or protein sequence (FASTA format or  
 simple format)  
 -o OUTPUT Output file (default: outfile.csv)  
 -t THRESHOLD Threshold (0-1, default: 0.5)  
 -j {1,2,3,4} Job type: 1-Prediction, 2-Protein Scanning, 3-Design,  
 4-Design all possible mutants  
 -m {1,2,3} Model selection: 1-ESM2-t30 (Peptides), 2-ET  
 (Proteins), 3-ET (Combined)  
 -d {1,2} Display: 1-Neurotoxic only, 2-All peptides (default)  
 -wd WORKING DIRECTORY  
 Working directory for saving results  
 -p POSITION Position to insert mutation (1-indexed)  
 -r RESIDUES Mutated residues (single/double letter amino acid codes)  
 -w {8-20} Window length (Protein Scan mode only, default: 12)

### 🔹 Input File Formats

NTxPred2 accepts two input formats:

1. **FASTA Format:**  
   >sequence\_id\_1  
   ADGHTYKLENMSTFGHQW  
   >sequence\_id\_2  
   PLKIHVSDWQERSAAS  
     
   (Example file: example.fasta)
2. **Simple Format:**  
   ADGHTYKLENMSTFGHQW  
   PLKIHVSDWQERSAAS  
     
   (Create a file with .seq extension, e.g., example.seq)

### 🔹 Output File

The output is a CSV file. Each row contains the sequence identifier and the predicted neurotoxicity score (probability).

## 🔍 Job Types and Features

NTxPred2 supports four job types:

|  |  |  |  |
| --- | --- | --- | --- |
| **Job Type** | **Description** | **Command Line Option (-j)** | **Additional Options** |
| Prediction | Predicts if a peptide/protein is neurotoxic. | 1 | -i, -o, -t, -m, -wd |
| Protein Scanning | Identifies neurotoxic regions within a protein. | 2 | -i, -o, -t, -m, -wd, -w, -d |
| Design | Generates single/di-amino acid mutants at a specific position. | 3 | -i, -o, -t, -m, -wd, -p, -r |
| Design All Possible Mutants | Generates and predicts all possible single amino acid mutants. | 4 | -i, -o, -t, -m, -wd, -p |

### Additional Options Explained

* -t THRESHOLD: The probability threshold for classifying a sequence as neurotoxic. Default is 0.5.
* -m MODEL: Selects the model (1, 2, or 3).
* -wd WORKING DIRECTORY: Specifies the directory for output files.
* -p POSITION: The position (1-indexed) for mutations in Design mode.
* -r RESIDUES: The amino acid residues to insert in Design mode. Use single-letter codes (e.g., -r AL) for single mutant, or double letter codes for dipeptide insertion.
* -w WINDOW: The window length for Protein Scanning mode (default: 12).
* -d DISPLAY: 1 for neurotoxic peptides only, 2 for all.

## 🎯 Example Usage Scenarios

1. **Basic Prediction:**  
   python ntxpred2.py -i input.fasta -o results.csv -j 1 -m 1 -wd ./results  
     
   Predicts the neurotoxicity of sequences in input.fasta using the peptide model, saves results to results.csv, in the folder ./results.
2. **Protein Scanning:**  
   python ntxpred2.py -i protein.fasta -o scan\_results.csv -j 2 -m 2 -w 15 -wd ./scan\_results  
     
   Scans protein.fasta for neurotoxic regions using a window of 15, saves results to scan\_results.csv.
3. **Design Mutant:**  
   python ntxpred2.py -i peptide.fasta -o mutant\_results.csv -j 3 -m 1 -p 5 -r "GA" -wd ./mutant\_results  
     
   Designs a mutant of the sequence in peptide.fasta by inserting Gly-Ala at position 5, and predicts its neurotoxicity.
4. **Design All Possible Mutants**  
   python ntxpred2.py -i protein.fasta -o all\_mutants.csv -j 4 -m 2 -p 10 -wd ./all\_mutants  
     
   Generates all possible single amino acid mutants at position 10 of the protein in protein.fasta, and predicts their neurotoxicity.

## 📦 Package Contents

The NTxPred2 package includes:

* INSTALLATION: Installation instructions (text file).
* LICENSE: License information (text file).
* README.md: Basic information (this file).
* ntxpred2.py: The main Python script.
* example.fasta: Example input file.
* Pre-trained model files (download separately).

## 🔗 Web Server

For online access, visit the NTxPred2 web server: <http://webs.iiitd.edu.in/raghava/ntxpred2>

## 🚀 Get Started

By following these instructions, you should be able to successfully install and use NTxPred2 for your neurotoxicity prediction needs.