

---

# **HDLTex Documentation**

*Release latest*

**Feb 19, 2019**



---

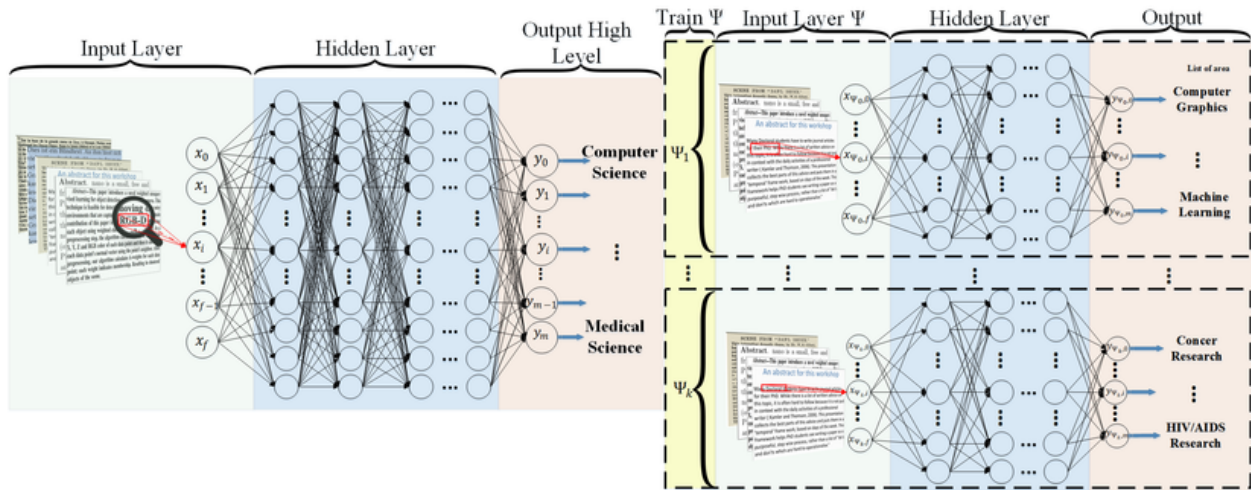
## Contents

---

<b>1 Documentation:</b>	<b>3</b>
<b>2 Installation</b>	<b>5</b>
2.1 Using pip . . . . .	5
2.2 Using git . . . . .	5
<b>3 Datasets for HDLTex:</b>	<b>7</b>
3.1 Requirements : . . . . .	7
<b>4 Feature Extraction:</b>	<b>9</b>
<b>5 Error and Comments:</b>	<b>11</b>
<b>6 Citation:</b>	<b>13</b>



Referenced paper : HDLTex: Hierarchical Deep Learning for Text Classification





# CHAPTER 1

---

## Documentation:

---

Increasingly large document collections require improved information processing methods for searching, retrieving, and organizing text. Central to these information processing methods is document classification, which has become an important application for supervised learning. Recently the performance of traditional supervised classifiers has degraded as the number of documents has increased. This is because along with growth in the number of documents has come an increase in the number of categories. This paper approaches this problem differently from current document classification methods that view the problem as multi-class classification. Instead we perform hierarchical classification using an approach we call Hierarchical Deep Learning for Text classification (HDLTex). HDLTex employs stacks of deep learning architectures to provide specialized understanding at each level of the document hierarchy.





### 2.1 Using pip

```
pip install HDLTex
```

### 2.2 Using git

```
git clone --recursive https://github.com/kk7nc/HDLTex.git
```

The primary requirements for this package are Python 3 with Tensorflow. The requirements.txt file contains a listing of the required Python packages; to install all requirements, run the following:

```
pip -r install requirements.txt
```

Or

```
pip3 install -r requirements.txt
```

Or:

```
conda install --file requirements.txt
```

If the above command does not work, use the following:

```
sudo -H pip install -r requirements.txt
```



---

## Datasets for HDLTex:

---

Linke of dataset:

Web of Science Dataset [WOS-11967](#)

```
This dataset contains 11,967 documents with 35 categories which include 7 parents_
↪categories.
```

Web of Science Dataset [WOS-46985](#)

```
This dataset contains 46,985 documents with 134 categories which include 7 parents_
↪categories.
```

Web of Science Dataset [WOS-5736](#)

```
This dataset contains 5,736 documents with 11 categories which include 3 parents_
↪categories.
```

## 3.1 Requirements :

General:

- Python 3.5 or later see [Instruction Documents](#)
- TensorFlow see [Instruction Documents](#).
- scikit-learn see [Instruction Documents](#)
- Keras see [Instruction Documents](#)
- scipy see [Instruction Documents](#)
- GPU
  - CUDA® Toolkit 8.0. For details, see [NVIDIA's documentation](#).
  - The [NVIDIA drivers](#) associated with CUDA Toolkit 8.0.

- cuDNN v6. For details, see [NVIDIA's documentation](#).
- GPU card with CUDA Compute Capability 3.0 or higher.
- The libcupti-dev library,
- To install this library, issue the following command:

```
$ sudo apt-get install libcupti-dev
```

## CHAPTER 4

---

### Feature Extraction:

---

#### Global Vectors for Word Representation (GLOVE)

For CNN **and** RNN you need to download **and** linked the folder location to GLOVE



## CHAPTER 5

---

Error and Comments:

---

Send an email to [kk7nc@virginia.edu](mailto:kk7nc@virginia.edu)





## CHAPTER 6

---

Citation:

---

```
@inproceedings{Kowsari2018HDLTex,  
author={Kowsari, Kamran and Brown, Donald E and Heidarysafa, Mojtaba and Meimandi,   
↪Kiana Jafari and Gerber, Matthew S and Barnes, Laura E},  
booktitle={2017 16th IEEE International Conference on Machine Learning and   
↪Applications (ICMLA)},  
title={HDLTex: Hierarchical Deep Learning for Text Classification},  
year={2017},  
pages={364-371},  
doi={10.1109/ICMLA.2017.0-134},  
month={Dec}  
}
```