Unit Testing for Tesseract

Requirements

Files and structure

```
├── langdata_lstm
│   ├── common.punc
│   ├── common.unicharambigs
│   ├── desired_bigrams.txt
│   ├── eng
│   │   ├── desired_characters
│   │   ├── eng.config
│   │   ├── eng.numbers
│   │   ├── eng.punc
│   │   ├── eng.singles_text
│   │   ├── eng.training_text
│   │   ├── eng.unicharambigs
│   │   ├── eng.wordlist
│   │   └── okfonts.txt
│   ├── extended
│   │   └── extended.config
│   ├── extendedhin
│   │   └── extendedhin.config
│   ├── font_properties
│   ├── forbidden_characters_default
│   ├── hin
│   │   ├── hin.config
│   │   ├── hin.numbers
│   │   ├── hin.punc
│   │   └── hin.wordlist
│   ├── kan
│   │   └── kan.config
│   ├── kor
│   │   └── kor.config
│   ├── osd
│   │   └── osd.unicharset
│   └── radical-stroke.txt
├── tessdata
│   ├── ara.traineddata
│   ├── chi_tra.traineddata
│   ├── eng.traineddata
│   ├── heb.traineddata
│   ├── hin.traineddata
│   ├── jpn.traineddata
│   ├── kmr.traineddata
│   ├── osd.traineddata
│   └── vie.traineddata
├── tessdata_best
```

```
│   ├── eng.traineddata
│   ├── fra.traineddata
│   ├── kmr.traineddata
│   └── osd.traineddata
├── tessdata_fast
│   ├── eng.traineddata
│   ├── kmr.traineddata
│   ├── osd.traineddata
│   └── script
│       └── Latin.traineddata
└── tesseract
    ...
    ├── test
    ├── unittest
    │   └── third_party/googletest
    └── VERSION
```

Fonts

- Microsoft fonts: arialbi.ttf, times.ttf, verdana.ttf - installation guide
- ae_Arab.ttf
- dejavu-fonts: DejaVuSans-ExtraLight.ttf
- Lohit-Hindi.ttf
- UnBatang.ttf

Run tests

To run the tests, do the following in tesseract folder

```
autoreconf -fiv
git submodule update --init
export TESSDATA_PREFIX=/prefix/to/path/to/tessdata
make check
```