Table 1 – Current and desired levels of precision/sensitivity

|  |  |  |  |
| --- | --- | --- | --- |
| **Training Data** | **Testing DATA** | **Precision/sensitivity**  **(current dataset & BANNER)** | **Desired improvement in precision/sensitivity (on current dataset &/or larger, newly generated dataset)** |
| Expert Dataset | XX% Expert Dataset | XX | XX |
| M-Turk Dataset | XX% Expert Dataset | XX | XX |
| M-Turk Dataset | XX% M-Turk Dataset | XX | XX |
| Expert Dataset &  M-Turk Dataset | XX% Expert Dataset | XX | XX |
| Expert Dataset &  M-Turk Dataset | XX% Expert Dataset & XX% M-Turk Dataset | XX | XX |

**Definitions & Descriptive statistics**:

1. **Expert Dataset**
   1. We have ~800 abstracts from published papers in the field(s) of […] annotated by **XX** experts. With a total of **XX** diseases identified overall (with a min of **XX** and max of **XX** per abstract).
2. **M-Turk Dataset**
   1. We have ~800 abstracts (the same of the Expert Dataset) annotated by **XX** Mechanical Turk workers (each annotating **XX** abstracts). With a total of **XX** diseases identified overall (with a min of **XX** and max of **XX** per abstract). With a fraction **XX** of diseases which were not identified by the experts, and a fraction **XX** of diseases which were identified by the experts but not by M-turkers.
3. **Diseases dictionary**
   1. A list of labels consisting of **XX** possible diseases that experts or M-turkers could have used in their annotation activity.

Below is a copy of the written instructions given to Experts & Mechanical Turk Workers respectively. […]