

CLAIMS

What is claimed is:

- 5 1. A method of federated machine learning using at least one processor, the method comprising:
transmitting a current global machine learning model to each of a plurality
of data sources;
receiving a plurality of training updates from the plurality of data sources,
10 respectively, each of the plurality of training updates being generated by the
respective data source in response to the global machine learning model received;
and
updating the current global machine learning model based on the plurality
of training updates received and a plurality of data quality parameters associated
15 with the plurality of data sources, respectively, to generate an updated global
machine learning model.
2. The method according to claim 1, wherein each of the plurality of training updates
is generated by the respective data source based on the global machine learning
20 model received and labelled data stored by the respective data source.
3. The method according to claim 2, wherein each of the plurality of training updates
comprises a difference between the current global machine learning model and a
local machine learning model trained by the respective data source based on the
25 current global machine learning model and labelled data stored by the respective
data source.
4. The method according to claim 1, wherein said updating the current global machine
learning model comprises determining a weighted average of the plurality of
30 training updates based on the plurality of data quality parameters associated with
the plurality of data sources, respectively.

5. The method according to claim 2, wherein the labelled data stored by the respective data source comprises features and labels, and the data quality parameter associated with the respective data source comprises at least one of a feature quality parameter associated with the features and a label quality parameter associated with the labels.

5
6. The method according to claim 5, wherein one or more of the plurality of data quality parameters are each based on at least one of a first data quality factor, a second data quality factor, and a third data quality factor, wherein the first data quality factor relates to a quality of the corresponding data source, the second data quality factor relates to a quality of labelled data stored by the corresponding data source, and the third data quality factor relates to a statistical derivation of data uncertainty.

10
7. The method according to claim 6, wherein the first data quality factor is based on at least one of a reputation level associated with the data source, a competence level of one or more data annotators of the labelled data stored by the corresponding data source, and a method value associated with a type of annotation method used to produce the labelled data stored by the corresponding data source, and wherein the features of the labelled data are related to images, and the second data quality factor is based on at least one of image acquisition characteristics and a level of image artifacts in the images.

15

20
8. The method according to claim 1, further comprising:
 - binning multiple data sources into a plurality of quality ranges; and
 - selecting the plurality of data sources from multiple data sources.

25
9. The method according to claim 1, wherein the plurality of data quality parameters are a plurality of data quality indices.
10. A server for federated machine learning comprising:
 - a memory; and

30

at least one processor communicatively coupled to the memory and configured to:

transmit a current global machine learning model to each of a plurality of data sources;

5 receive a plurality of training updates from the plurality of data sources, respectively, each of the plurality of training updates being generated by the respective data source in response to the global machine learning model received; and

10 update the current global machine learning model based on the plurality of training updates received and a plurality of data quality parameters associated with the plurality of data sources, respectively, to generate an updated global machine learning model.

11. The server according to claim 10, wherein each of the plurality of training updates
15 is generated by the respective data source based on the global machine learning model received and labelled data stored by the respective data source.

12. The server according to claim 11, wherein each of the plurality of training updates
20 comprises a difference between the current global machine learning model and a local machine learning model trained by the respective data source based on the current global machine learning model and labelled data stored by the respective data source.

13. The server according to claim 10, wherein said update the current global machine
25 learning model comprises determining a weighted average of the plurality of training updates based on the plurality of data quality parameters associated with the plurality of data sources, respectively.

14. The server according to claim 11, wherein the labelled data stored by the respective
30 data source comprises features and labels, and the data quality parameter associated

with the respective data source comprises at least one of a feature quality parameter associated with the features and a label quality parameter associated with the labels

15. The server according to claim 14, wherein one or more of the plurality of data
5 quality parameters are each based on at least one of a first data quality factor, a second data quality factor, and a third data quality factor, wherein the first data quality factor relates to a quality of the corresponding data source, the second data quality factor relates to a quality of labelled data stored by the corresponding data source, and the third data quality factor relates to a statistical derivation of data
10 uncertainty.
16. The server according to claim 14, wherein the first data quality factor is based on at least one of a reputation level associated with the data source, a competence level of one or more data annotators of the labelled data stored by the corresponding data
15 source, and a method value associated with a type of annotation method used to produce the labelled data stored by the corresponding data source, and wherein the features of the labelled data are related to images, and the second data quality factor is based on at least one of image acquisition characteristics and a level of image artifacts in the images.
- 20 17. The server according to claim 10, wherein the at least one processor is further configured to:
- bin multiple data sources into a plurality of quality ranges; and
select the plurality of data sources from multiple data sources.
- 25 18. The server according to claim 10, wherein the plurality of data quality parameters are a plurality of data quality indices.
- 30 19. A computer program product, embodied in one or more non-transitory computer-readable storage mediums, comprising instructions executable by at least one

processor to perform a method of federated machine learning, the method comprising:

transmitting a current global machine learning model to each of a plurality of data sources;

5 receiving a plurality of training updates from the plurality of data sources, respectively, each of the plurality of training updates being generated by the respective data source in response to the global machine learning model received; and

10 updating the current global machine learning model based on the plurality of training updates received and a plurality of data quality parameters associated with the plurality of data sources, respectively, to generate an updated global machine learning model.