CLAIMS

Claim 1. A computer-implemented method of identifying unreliable clusters to improve accuracy and efficiency of base calling, the method including:

accessing per-cycle cluster data for a plurality of clusters and for a first subset of sequencing cycles of a sequencing run;

base calling of a multi-pixel image through a neural network-based base caller each cluster in the plurality of clusters at each sequencing cycle in the first subset of sequencing cycles, including

processing the per-cycle cluster data and generating intermediate representations of the per-cycle cluster data, and

processing the intermediate representations though a normalization function of an output layer of the neural network-based base caller and producing a per-cluster, per-cycle probability quadruple for each cluster and for each sequencing cycle, wherein a particular per-cluster, per-cycle probability quadruple identifies probabilities of a base incorporated in a particular cluster at a particular sequencing cycle being A, C, T, and G;

using the identified probabilities to determine a filter value as a function of a highest probability that determines the base call for each per-cluster, per-cycle probability quadruple, thereby generating a sequence of filter values for each cluster;

identifying as unreliable clusters those clusters in the plurality of clusters whose sequences of filter values contain at least “N” number of filter values below a threshold “M”; and bypassing base calling of the unreliable clusters at a remainder of sequencing cycles of the sequencing run, thereby base calling, at the remainder of sequencing cycles, only those clusters in the plurality of clusters that are not identified as the unreliable clusters.

Claim 2. The computer-implemented method of claim 1, wherein the filter value for a per-cluster, per-cycle probability quadruple is determined based on an arithmetic operation involving one or more of the probabilities.

Claim 3. The computer-implemented method of claim 2, wherein the arithmetic operation is subtraction.

Claim 4. The computer-implemented method of claims 1-3, wherein the filter value for the per-cluster, per-cycle probability quadruple is determined by subtracting a second highest one of the probabilities from a highest one of the probabilities.

Claim 5. The computer-implemented method of claims 2-4, wherein the arithmetic operation is division.

Claim 6. The computer-implemented method of claims 1-3, wherein the filter value for the per-cluster, per-cycle probability quadruple is determined as a ratio of the highest one of the probabilities to the second highest one of the probabilities.

Claim 7. The computer-implemented method of claims 2, wherein the arithmetic operation is addition.

Claim 8. The computer-implemented method of claims 2, wherein the arithmetic operation is multiplication.

Claim 9. The computer-implemented method of claims 1-8, wherein the “N” ranges from 1 to 5.

Claim 10. The computer-implemented method of claims 1-9, wherein the “M” ranges from 0.5 to 0.99.

Claim 11. The computer-implemented method of claims 1-10, wherein the first subset includes 1 to 25 sequencing cycles of the sequencing run.

Claim 12. The computer-implemented method of claims 1-11, wherein the first subset includes 1 to 50 sequencing cycles of the sequencing run.

Claim 13. The computer-implemented method of claims 1-12, wherein the output layer is a softmax layer and the probabilities in the per-cluster, per-cycle probability quadruple are exponentially normalized classification scores that sum to unity.

Claim 14. The computer-implemented method of claims 1-13, wherein the unreliable clusters are indicative of empty, polyclonal, and dim wells on a patterned flow cell.

Claim 15. A system for improving accuracy and efficiency of neural network-based base calling, the system comprising:

memory storing, for a plurality of clusters, initial cluster data as a multi-pixel image for initial sequencing cycles of a sequencing run and remainder cluster data for remainder sequencing cycles of the sequencing run;

a host processor having access to the memory and configured to execute a detection and filtering logic to identify unreliable clusters;

a configurable processor having access to the memory and configured to execute a neural network to produce base call classification scores; and

a data flow logic having access to the memory, the host processor, and the configurable processor and configured

to provide the initial cluster data to the neural network and cause the neural network to generate initial intermediate representations from the initial cluster data and produce initial base call classification scores for the plurality of clusters and for the initial sequencing cycles by processing the initial intermediate representations through a normalization function of an output layer of the neural network,

to provide the initial base call classification scores to the detection and filtering logic and cause the detection and filtering logic to use the initial base call classification scores to generate filter values as a function of a highest classification score that determines a base call and to identify unreliable clusters in the plurality of clusters based on the generated filter values,

to provide the remainder cluster data to the neural network and cause the neural network to generate remainder intermediate representations from the remainder cluster data, and

to provide data identifying the unreliable clusters to the configurable processor and cause the configurable processor to generate reliable remainder intermediate representations by removing, from the remainder intermediate representations, those portions that result from portions of the remainder cluster data that represent the unreliable clusters.

Claim 16. The system of claim 15, wherein the system is further configured to produce the initial base call classification scores for the plurality of clusters and for the initial sequencing cycles by producing a per-cluster, per-cycle probability quadruple for each cluster and for each sequencing cycle, wherein a particular per-cluster, per-cycle probability

quadruple identifies probabilities of a base incorporated in a particular cluster at a particular sequencing cycle being A, C, T, and G.

Claim 17. The system of claim 16, wherein a filter value for a per-cluster, per-cycle probability quadruple is determined based on an arithmetic operation involving one or more of the probabilities.

Claim 18. The system of claim 17, wherein the arithmetic operation is subtraction, division, addition, or multiplication.

Claim 19. The system of claim 17, wherein the filter value for the per-cluster, per-cycle probability quadruple is determined by subtracting a second highest one of the probabilities from a highest one of the probabilities.

Claim 20. The system of claim 17, wherein the filter value for the per-cluster, per-cycle probability quadruple is determined as a ratio of the highest one of the probabilities to a second highest one of the probabilities.