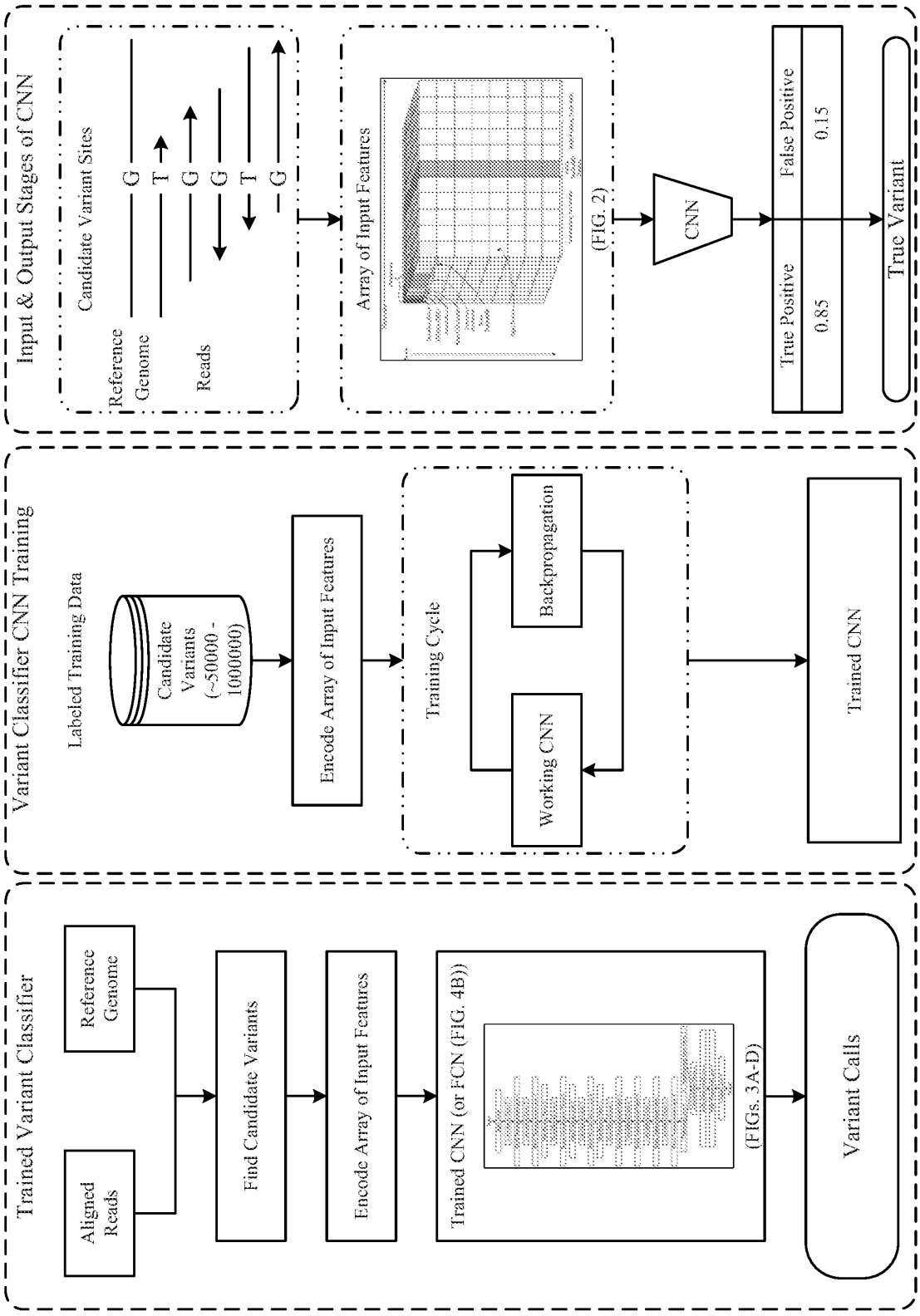

FIG. 1A

FIG. 1B

FIG. 1C

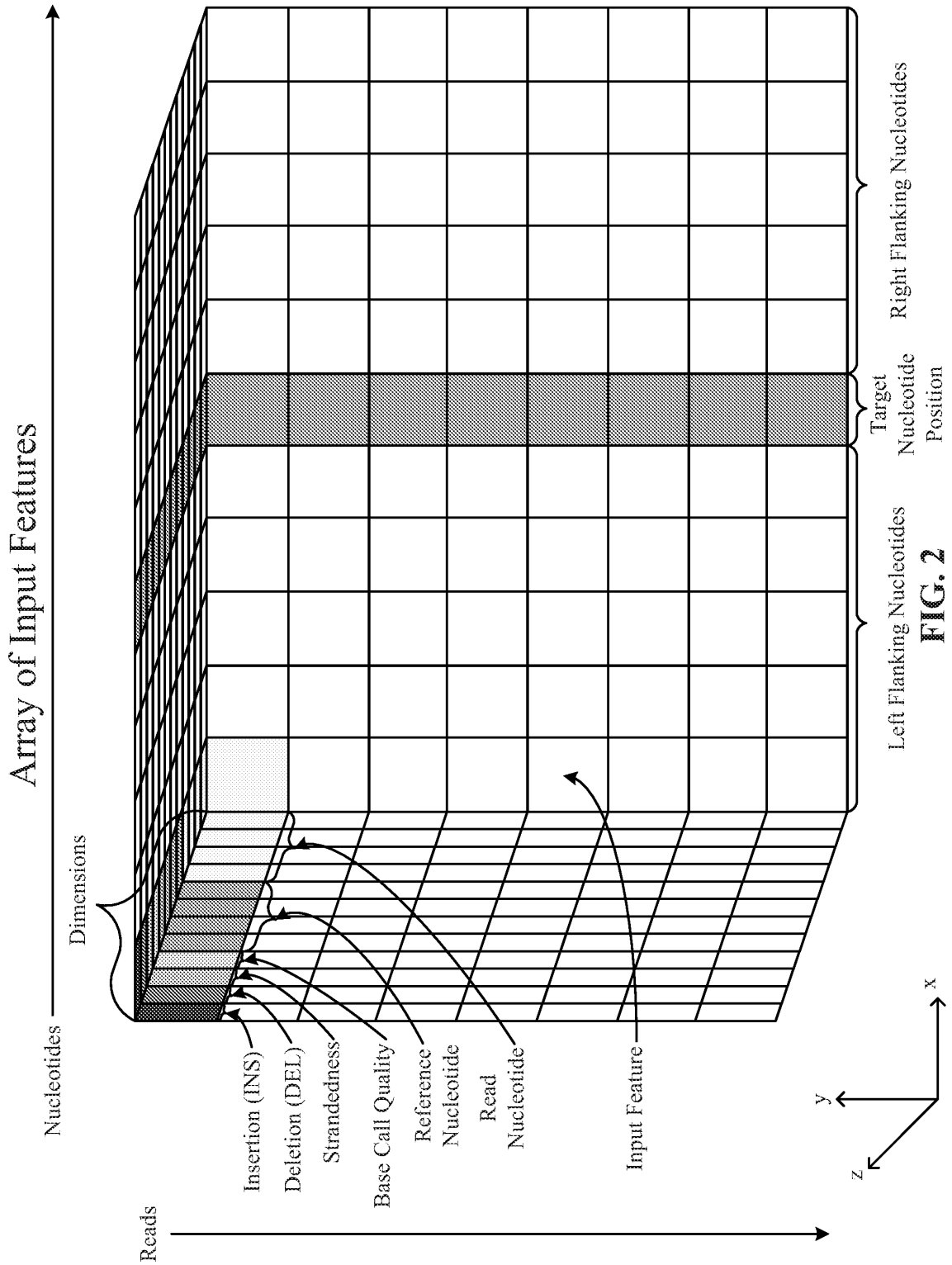
Array of Input Features
Nucleotides
Dimensions
Reads
Insertion (INS)
Deletion (DEL)
Strandedness
Base Call Quality
Reference Nucleotide
Read Nucleotide
Input Feature
Left Flanking Nucleotides
Target Nucleotide Position
Right Flanking Nucleotides
x
y
z

FIG. 2

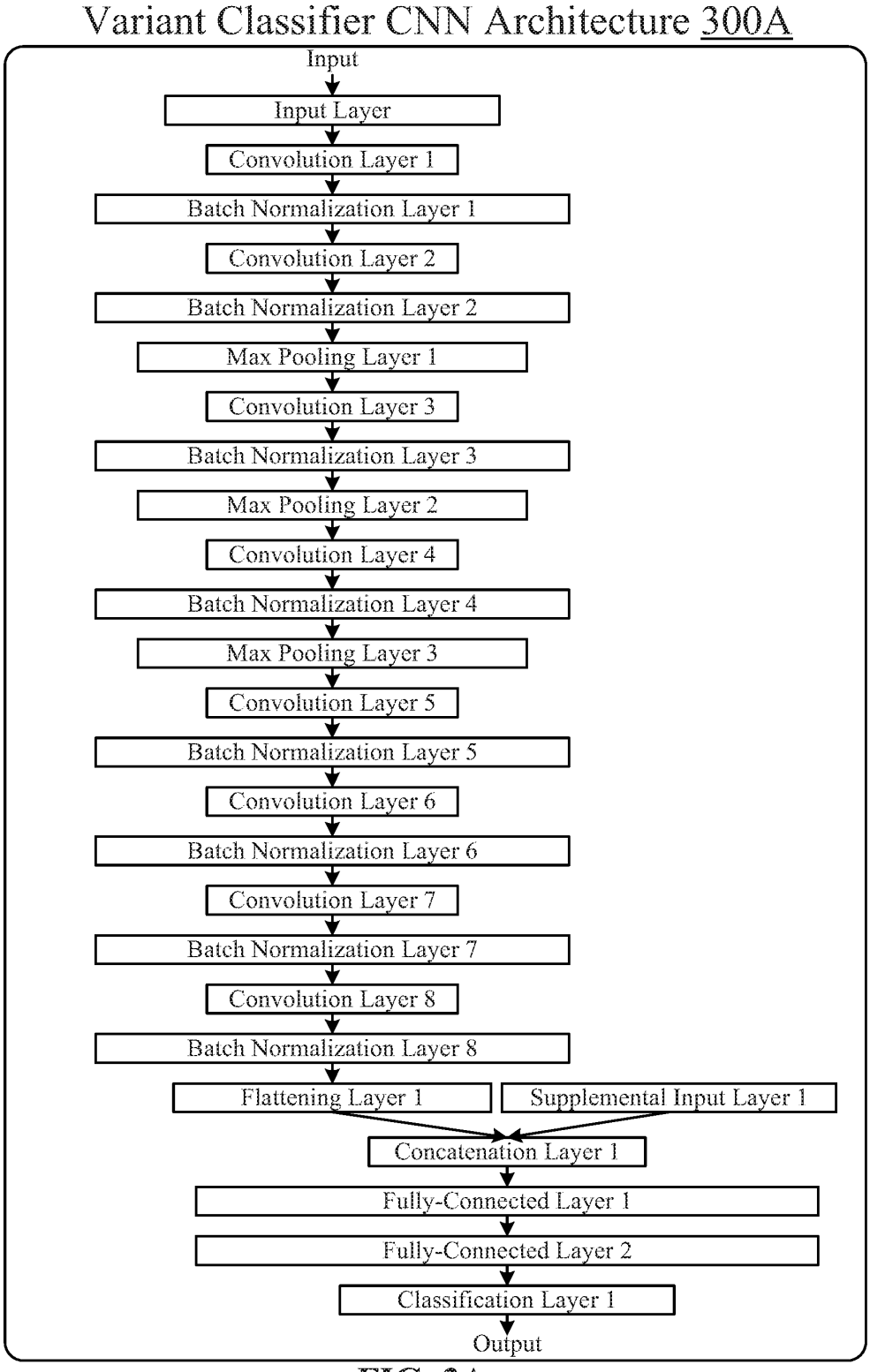
Variant Classifier CNN Architecture 300A
Input
Input Layer
Convolution Layer 1
Batch Normalization Layer 1
Convolution Layer 2
Batch Normalization Layer 2
Max Pooling Layer 1
Convolution Layer 3
Batch Normalization Layer 3
Max Pooling Layer 2
Convolution Layer 4
Batch Normalization Layer 4
Max Pooling Layer 3
Convolution Layer 5
Batch Normalization Layer 5
Convolution Layer 6
Batch Normalization Layer 6
Convolution Layer 7
Batch Normalization Layer 7
Convolution Layer 8
Batch Normalization Layer 8
Flattening Layer 1
Supplemental Input Layer 1
Concatenation Layer 1
Fully-Connected Layer 1
Fully-Connected Layer 2
Classification Layer 1
Output

FIG. 3A

Variant Classifier CNN Architecture 300B

- 220x100x10 Input comprising the following channels:
  - One-hot encode ACGT of the read (4 bits)
  - One-hot encode ACGT of the reference (4 bits)
  - Quality score (1 bit)
  - Strand (1 bit)
- CONV 5x5 and Maxpool (2:1, 2:1) → 108x48x32
- CONV 5x5 and Maxpool (2:1, 2:1) → 52x22x32
- CONV 5x5 and Maxpool (2:1, 2:1) → 25x9x32
- CONV 5x5, no Maxpool → 20x5x32
- CONV 5x5, no Maxpool → 16x1x32
- CONV 5x5, no Maxpool → 11x1x32
- CONV 5x5, no Maxpool → 7x1x32
- → Flatten to 224 Units + Concatenate 23 EVS Features → 247 Units
- Fully-Connected 256 Units
- → Output

FIG. 3B

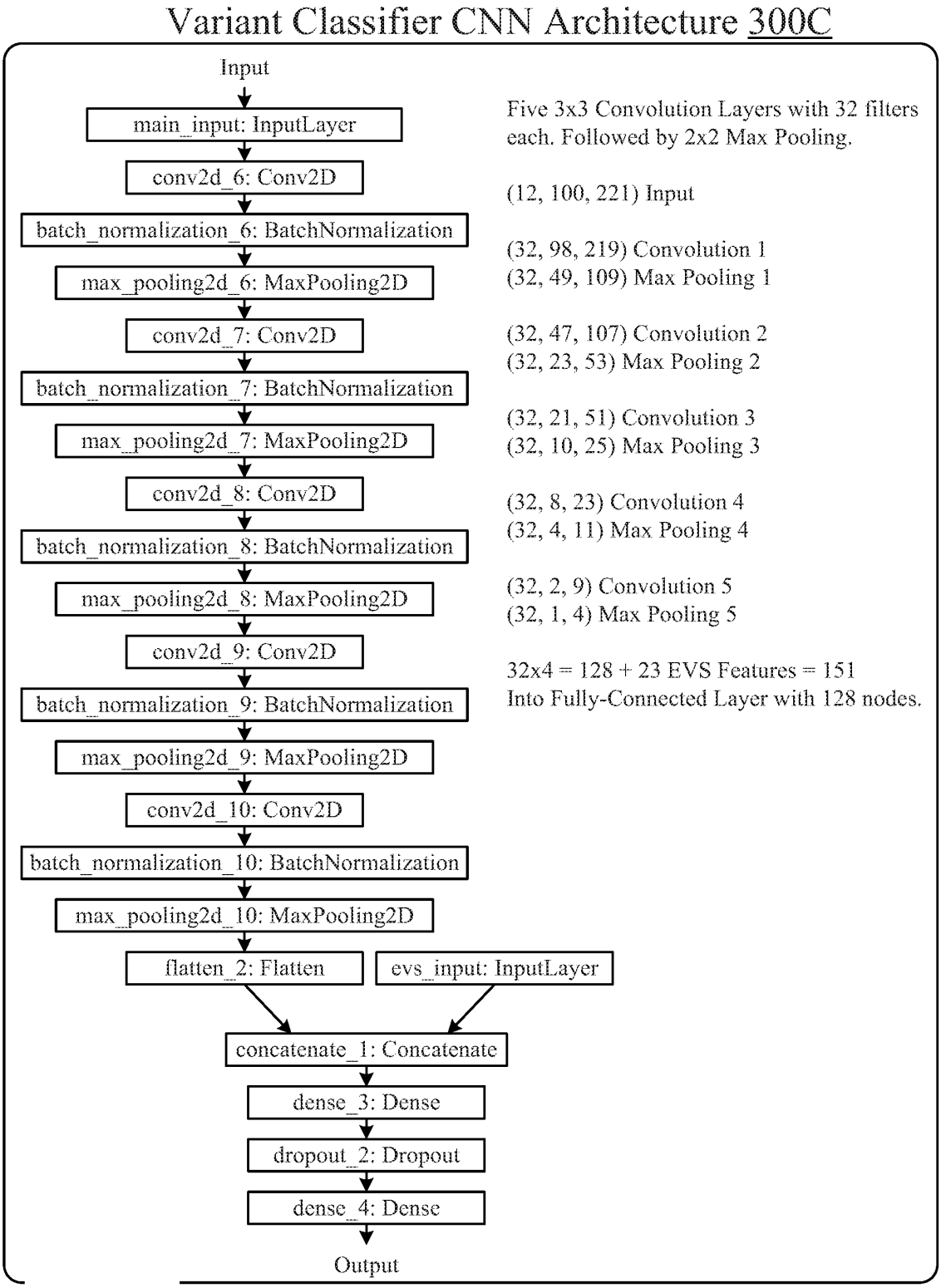
Variant Classifier CNN Architecture 300C
Input
main_input: InputLayer
conv2d_6: Conv2D
batch_normalization_6: BatchNormalization
max_pooling2d_6: MaxPooling2D
conv2d_7: Conv2D
batch_normalization_7: BatchNormalization
max_pooling2d_7: MaxPooling2D
conv2d_8: Conv2D
batch_normalization_8: BatchNormalization
max_pooling2d_8: MaxPooling2D
conv2d_9: Conv2D
batch_normalization_9: BatchNormalization
max_pooling2d_9: MaxPooling2D
conv2d_10: Conv2D
batch_normalization_10: BatchNormalization
max_pooling2d_10: MaxPooling2D
flatten_2: Flatten
evs_input: InputLayer
concatenate_1: Concatenate
dense_3: Dense
dropout_2: Dropout
dense_4: Dense
Output
Five 3x3 Convolution Layers with 32 filters each. Followed by 2x2 Max Pooling.
(12, 100, 221) Input
(32, 98, 219) Convolution 1
(32, 49, 109) Max Pooling 1
(32, 47, 107) Convolution 2
(32, 23, 53) Max Pooling 2
(32, 21, 51) Convolution 3
(32, 10, 25) Max Pooling 3
(32, 8, 23) Convolution 4
(32, 4, 11) Max Pooling 4
(32, 2, 9) Convolution 5
(32, 1, 4) Max Pooling 5
32x4 = 128 + 23 EVS Features = 151
Into Fully-Connected Layer with 128 nodes.

FIG. 3C

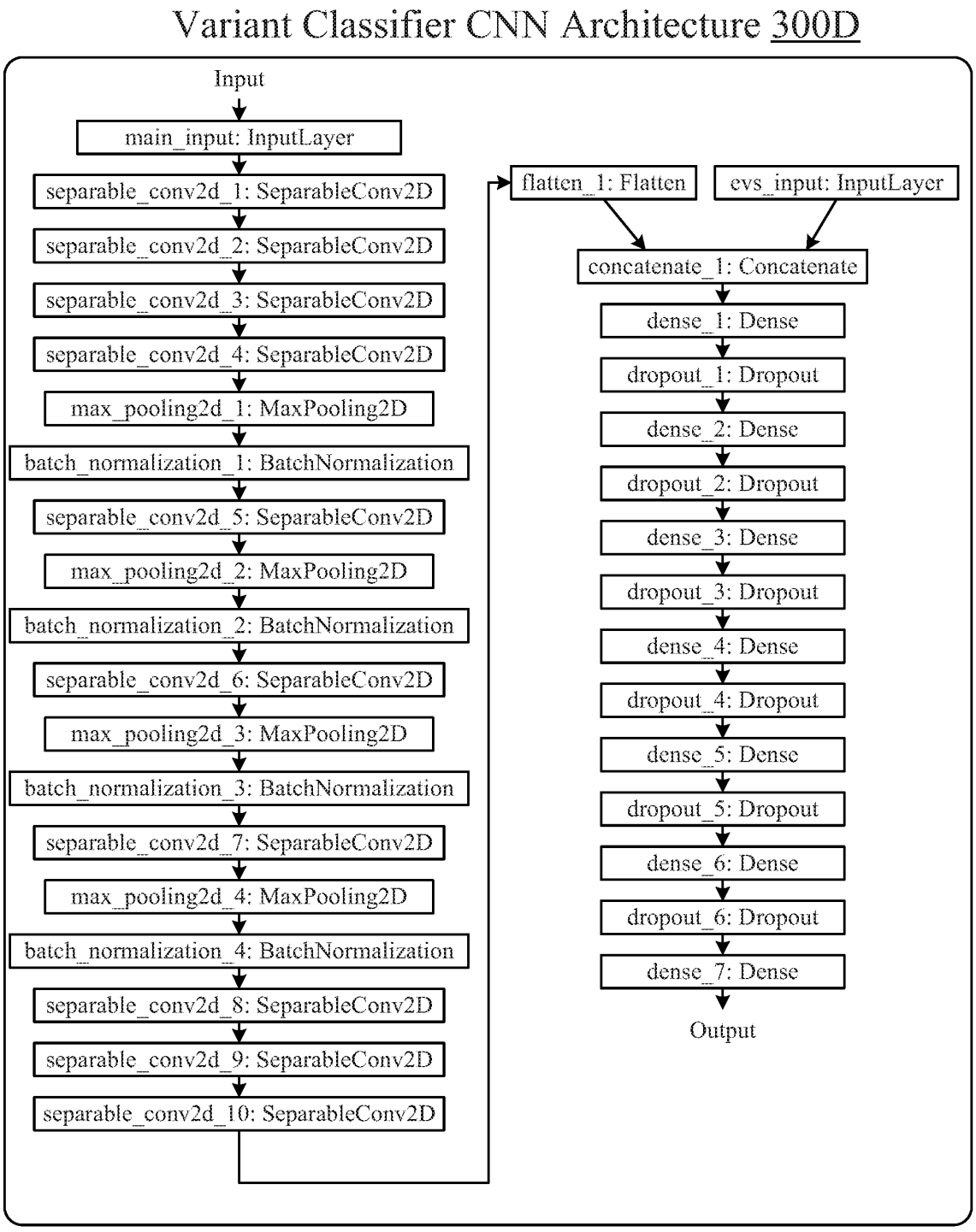
Variant Classifier CNN Architecture 300D
Input
main_input: InputLayer
separable_conv2d_1: SeparableConv2D
separable_conv2d_2: SeparableConv2D
separable_conv2d_3: SeparableConv2D
separable_conv2d_4: SeparableConv2D
max_pooling2d_1: MaxPooling2D
batch_normalization_1: BatchNormalization
separable_conv2d_5: SeparableConv2D
max_pooling2d_2: MaxPooling2D
batch_normalization_2: BatchNormalization
separable_conv2d_6: SeparableConv2D
max_pooling2d_3: MaxPooling2D
batch_normalization_3: BatchNormalization
separable_conv2d_7: SeparableConv2D
max_pooling2d_4: MaxPooling2D
batch_normalization_4: BatchNormalization
separable_conv2d_8: SeparableConv2D
separable_conv2d_9: SeparableConv2D
separable_conv2d_10: SeparableConv2D
flatten_1: Flatten
evs_input: InputLayer
concatenate_1: Concatenate
dense_1: Dense
dropout_1: Dropout
dense_2: Dense
dropout_2: Dropout
dense_3: Dense
dropout_3: Dropout
dense_4: Dense
dropout_4: Dropout
dense_5: Dense
dropout_5: Dropout
dense_6: Dense
dropout_6: Dropout
dense_7: Dense
Output

FIG. 3D

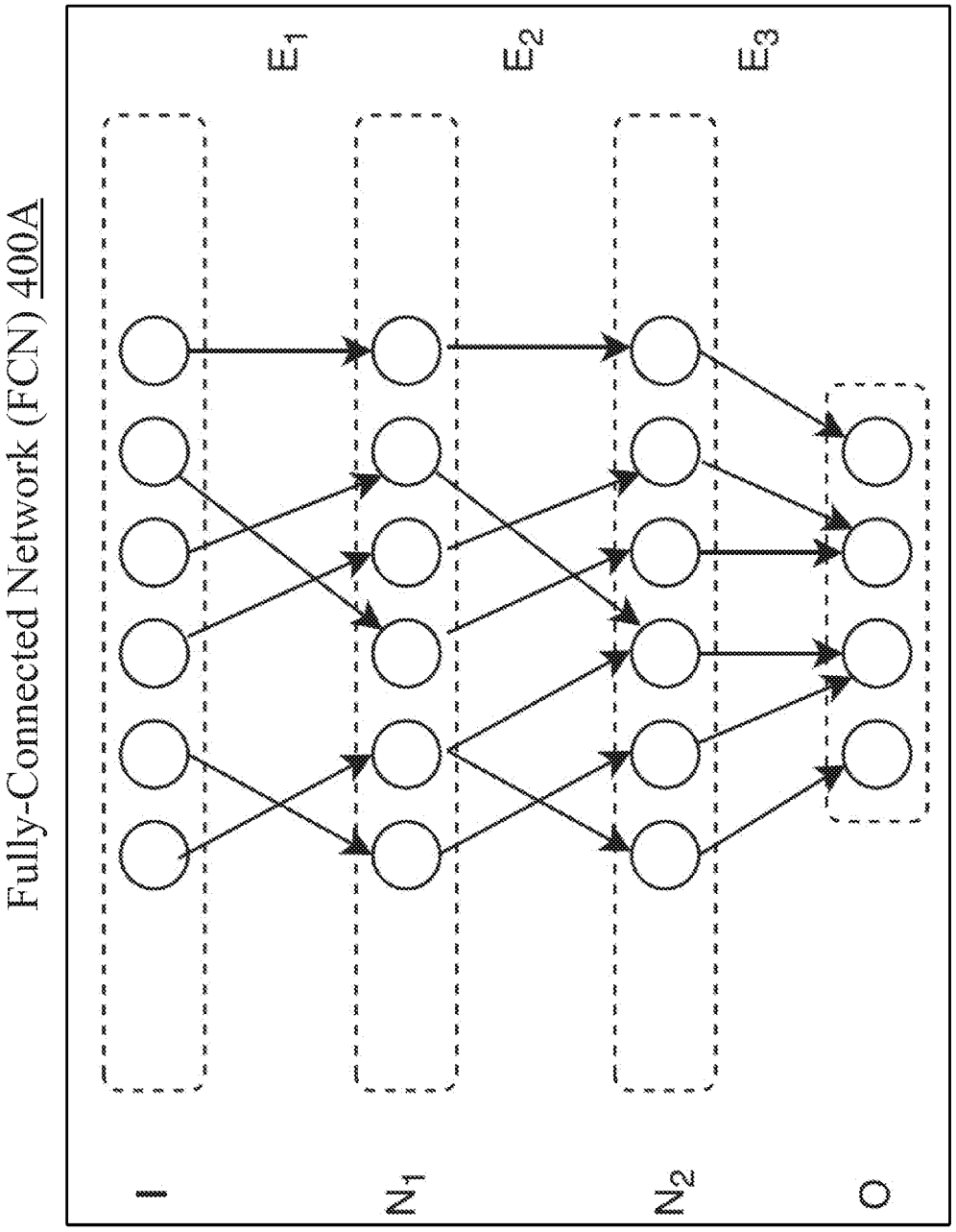

FIG. 4A

Variant Classifier FCN Architecture 400B

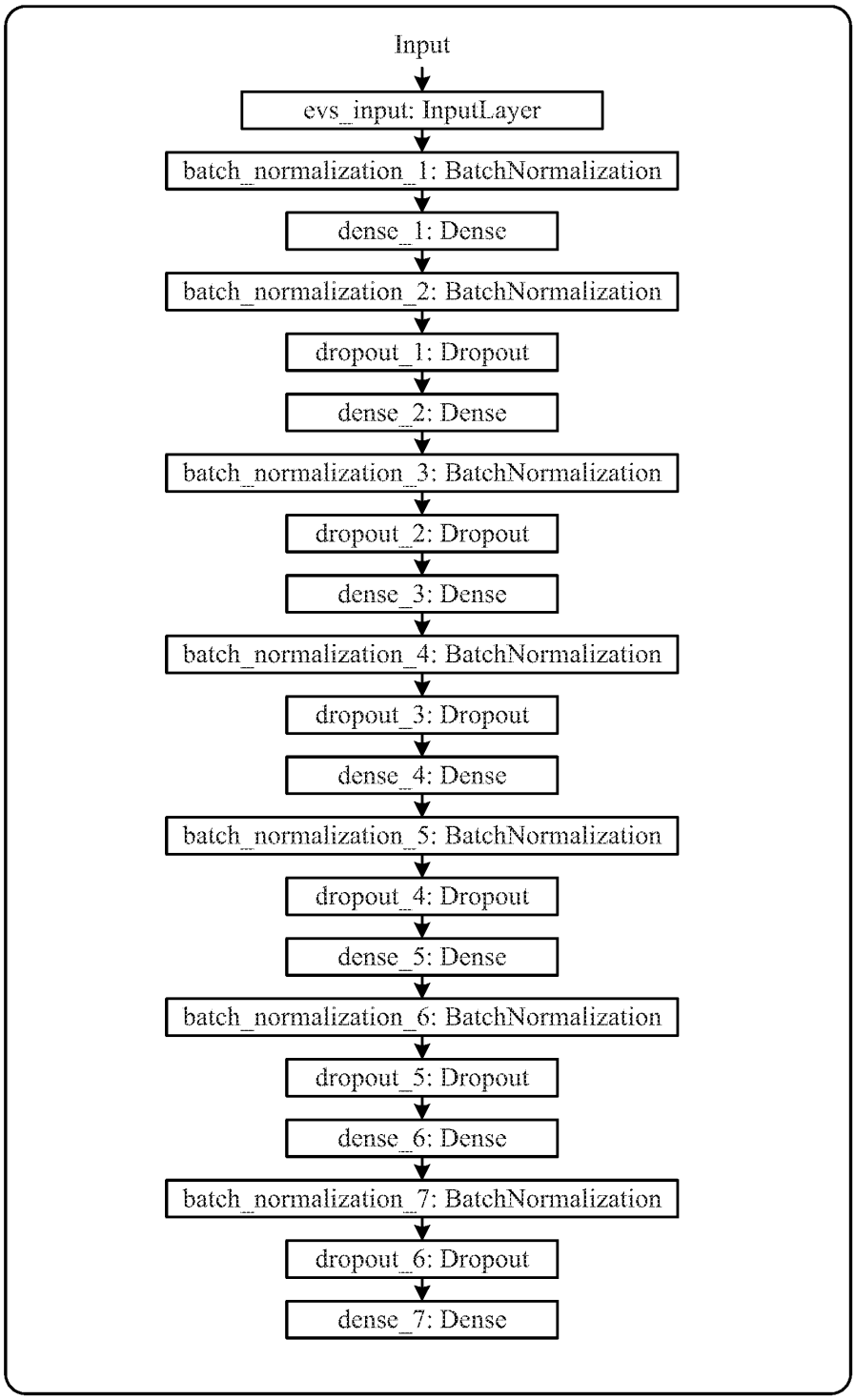

FIG. 4B

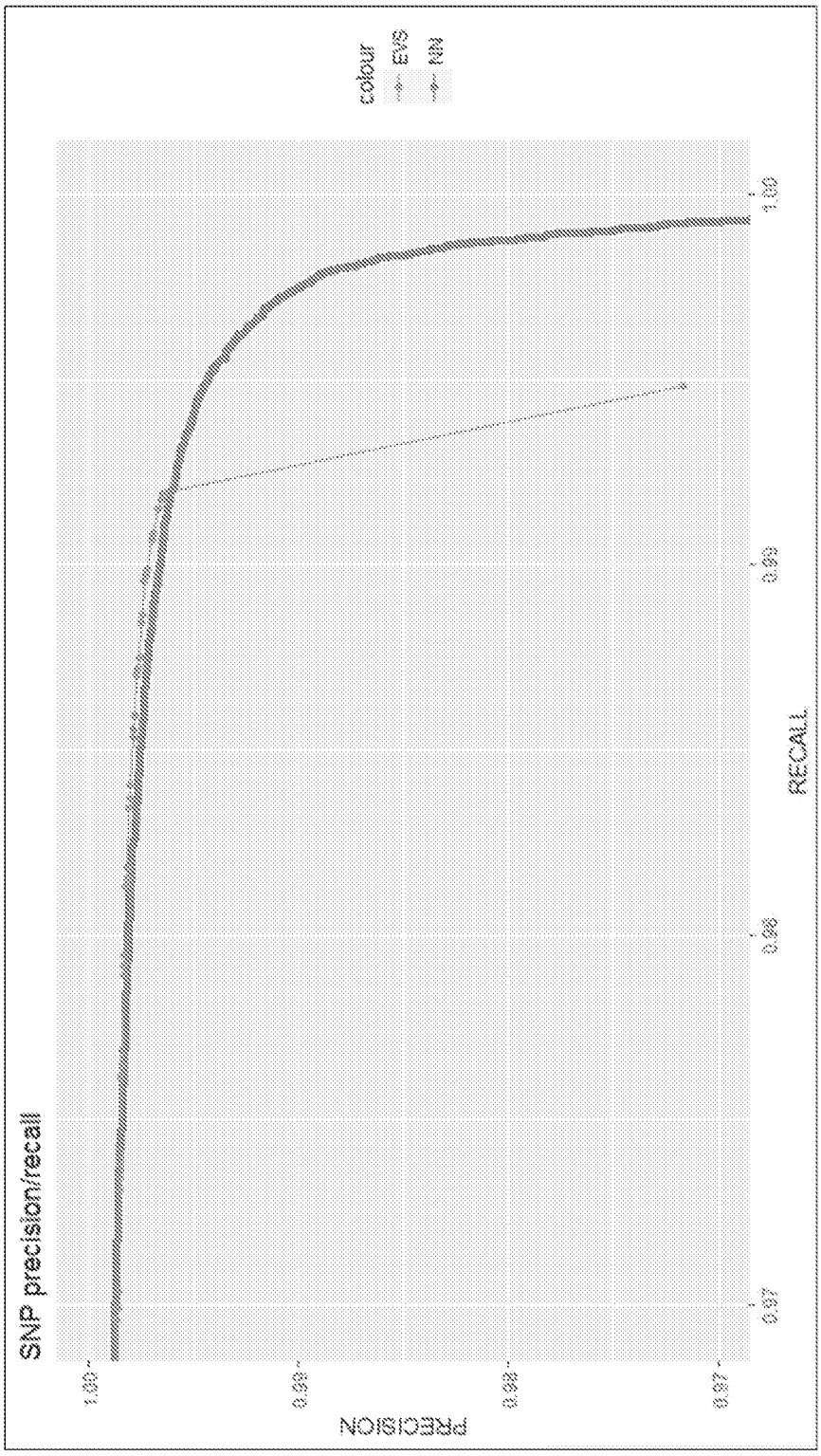
SNP precision/recall
colour
EVS
NN
PRECISION
RECALL
1.00
0.99
0.98
0.97
0.97
0.98
0.99
1.00

FIG. 5

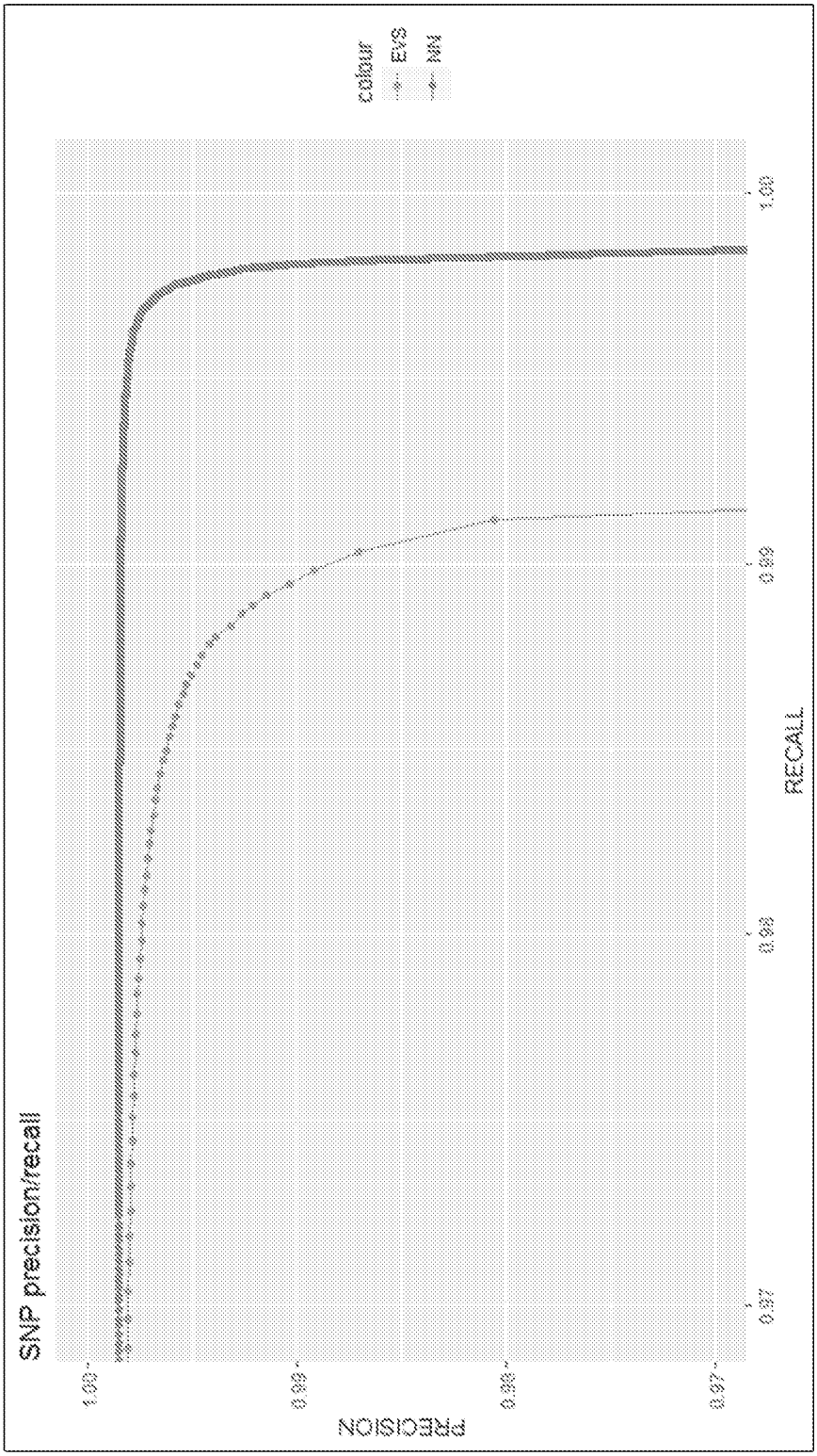
SNP precision/recall
colour
EXS
MN
PRECISION
1.00
0.99
0.98
0.97
RECALL
0.97
0.98
0.99
1.00

FIG. 6

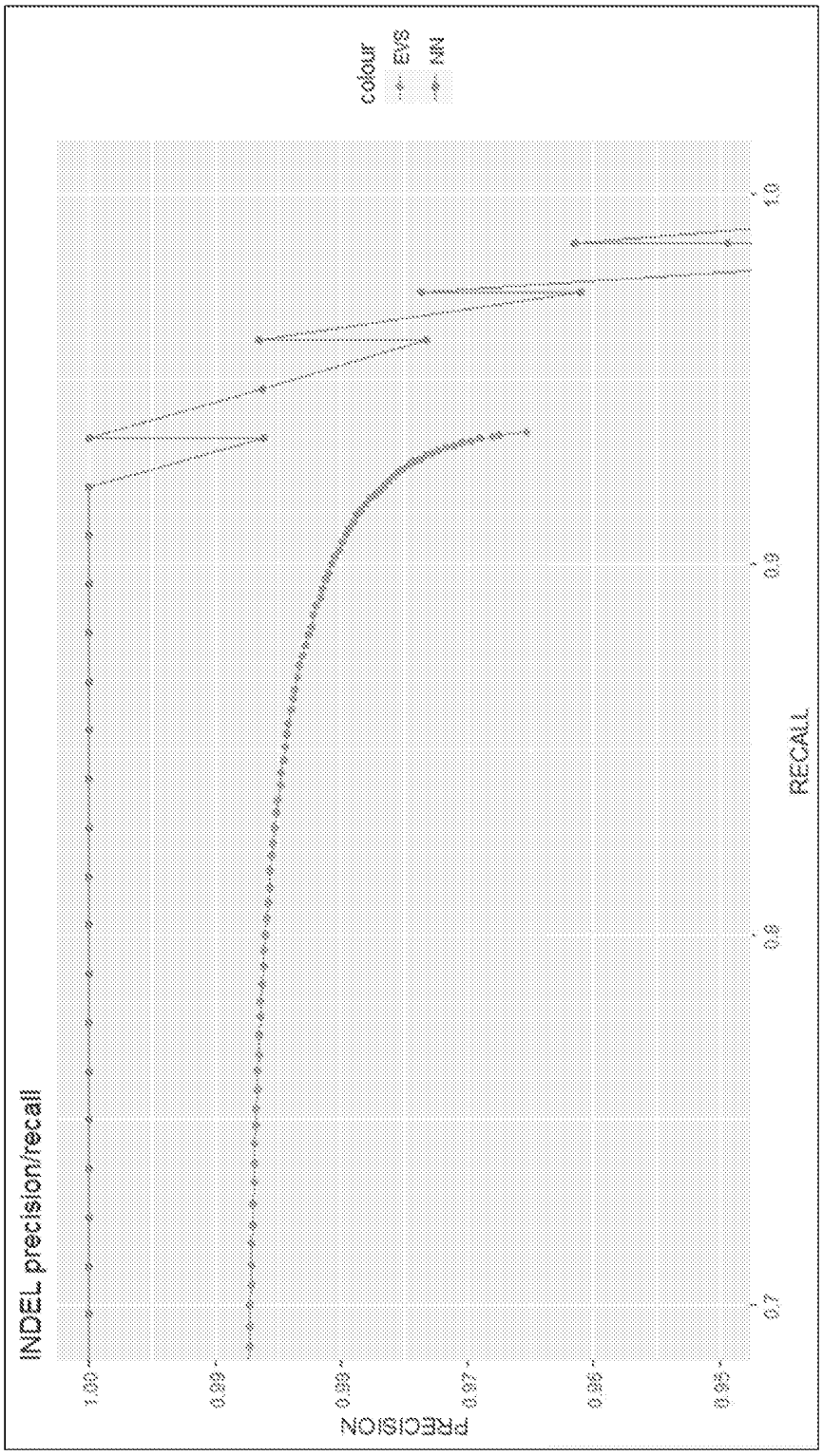
INDEL precision/recall
colour
EVS
MN
PRECISION
1.00
0.99
0.98
0.97
0.96
0.95
RECALL
0.7
0.8
0.9
1.0

FIG. 7

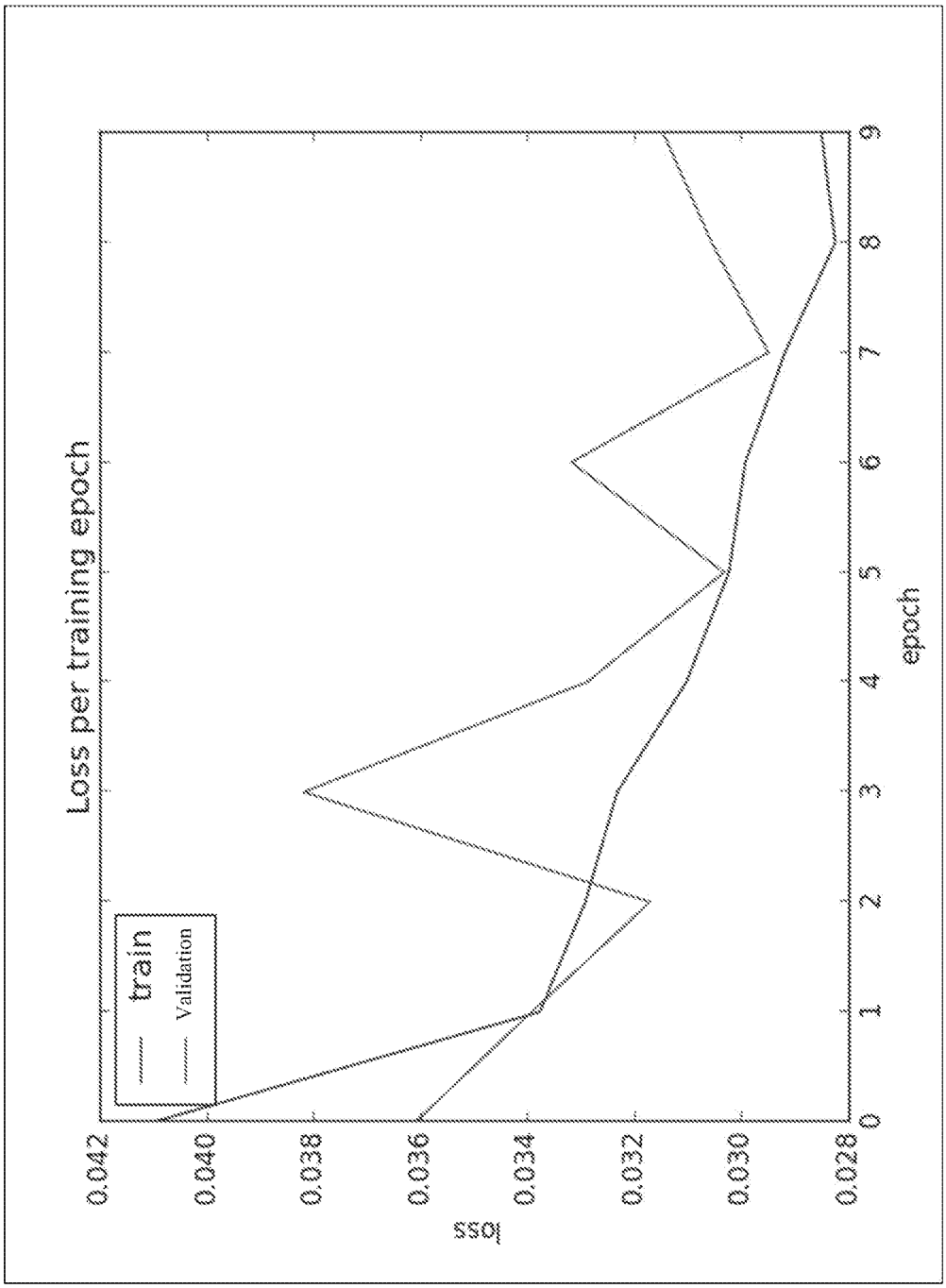
Loss per training epoch
train
Validation
loss
epoch
0.042
0.040
0.038
0.036
0.034
0.032
0.030
0.028
0
1
2
3
4
5
6
7
8
9

FIG. 8

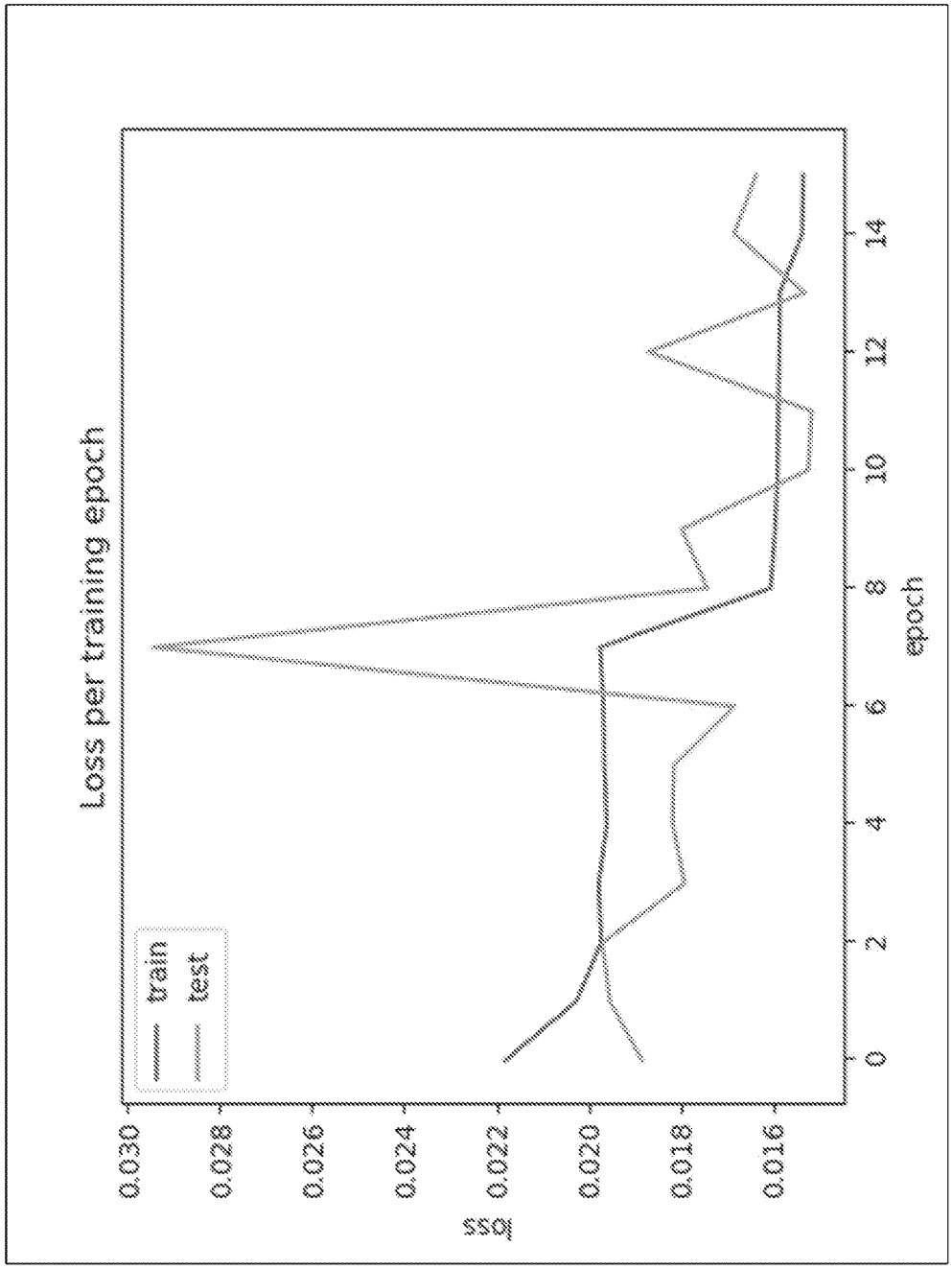
Loss per training epoch
train
test
loss
epoch
0.030
0.028
0.026
0.024
0.022
0.020
0.018
0.016
0
2
4
6
8
10
12
14

FIG. 9

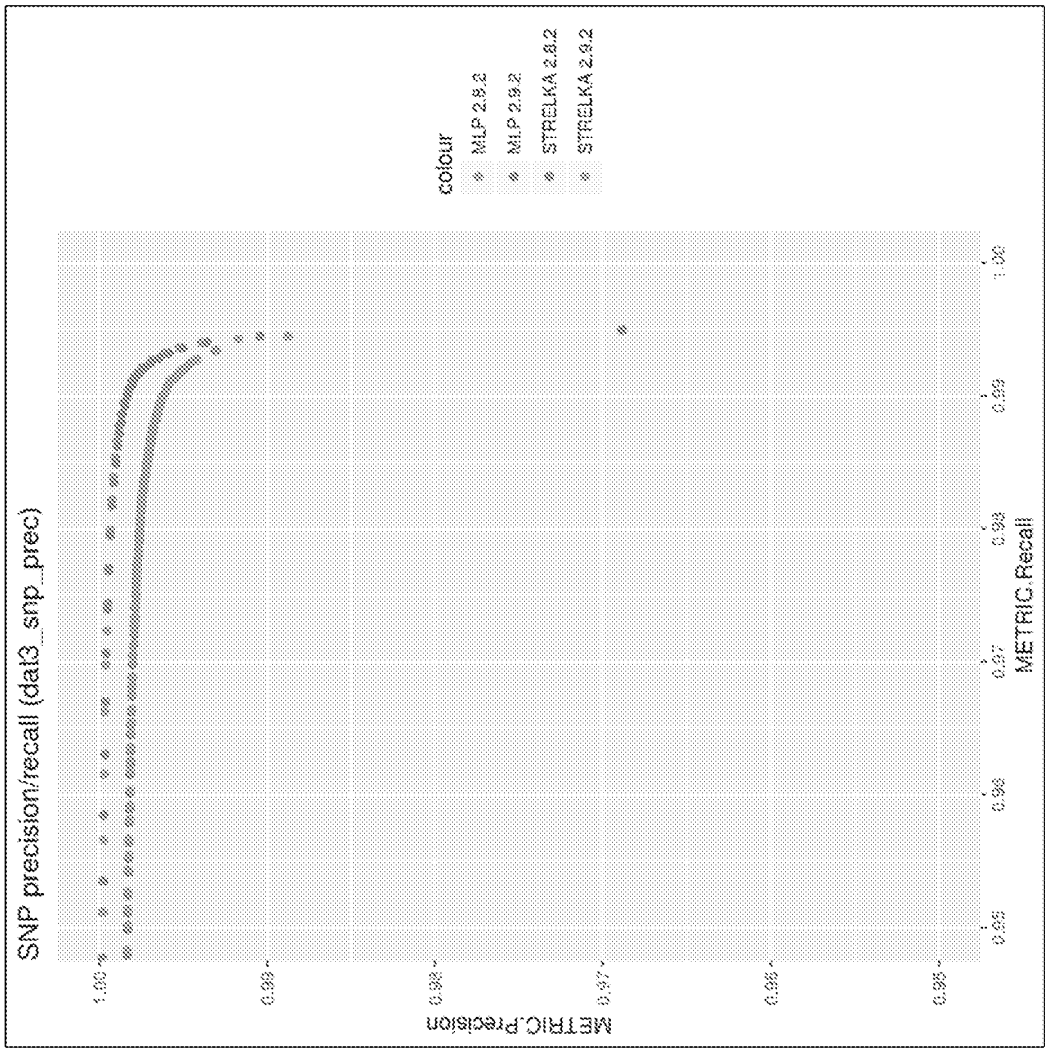
SNP precision/recall (dat3_snp_prec)
colour
MLP 2.8.2
MLP 2.9.2
STRELKA 2.8.2
STRELKA 2.9.2
METRIC.Precision
METRIC.Recall
1.00
0.99
0.98
0.97
0.96
0.95
0.95
0.96
0.97
0.98
0.99
1.00

FIG. 10

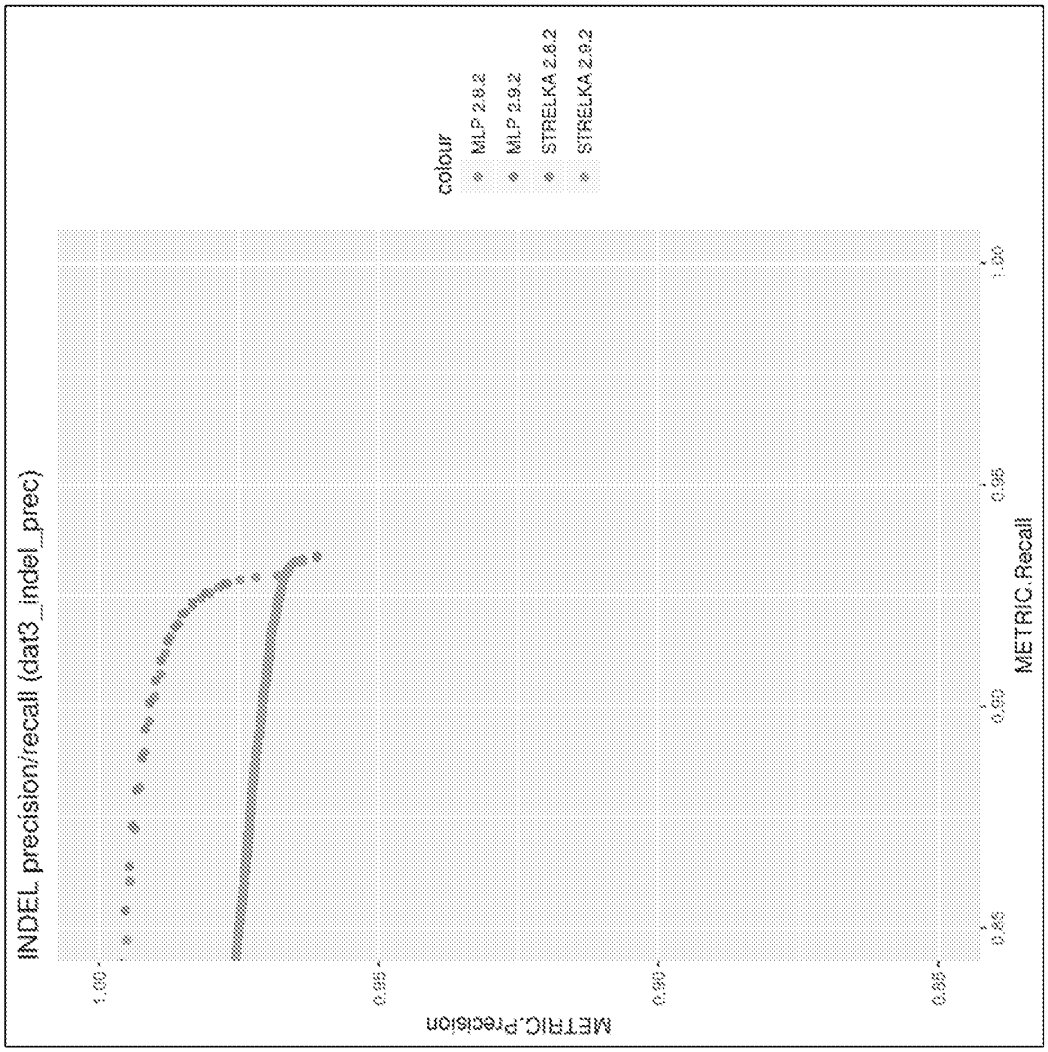
INDEL precision/recall (dat3_indel_prec)
colour
MLP 2.9.2
MLP 2.9.2
STRELKA 2.8.2
STRELKA 2.9.2
METRIC.Precision
METRIC.Recall
1.00
0.95
0.90
0.85
0.85
0.90
0.95
1.00

FIG. 11

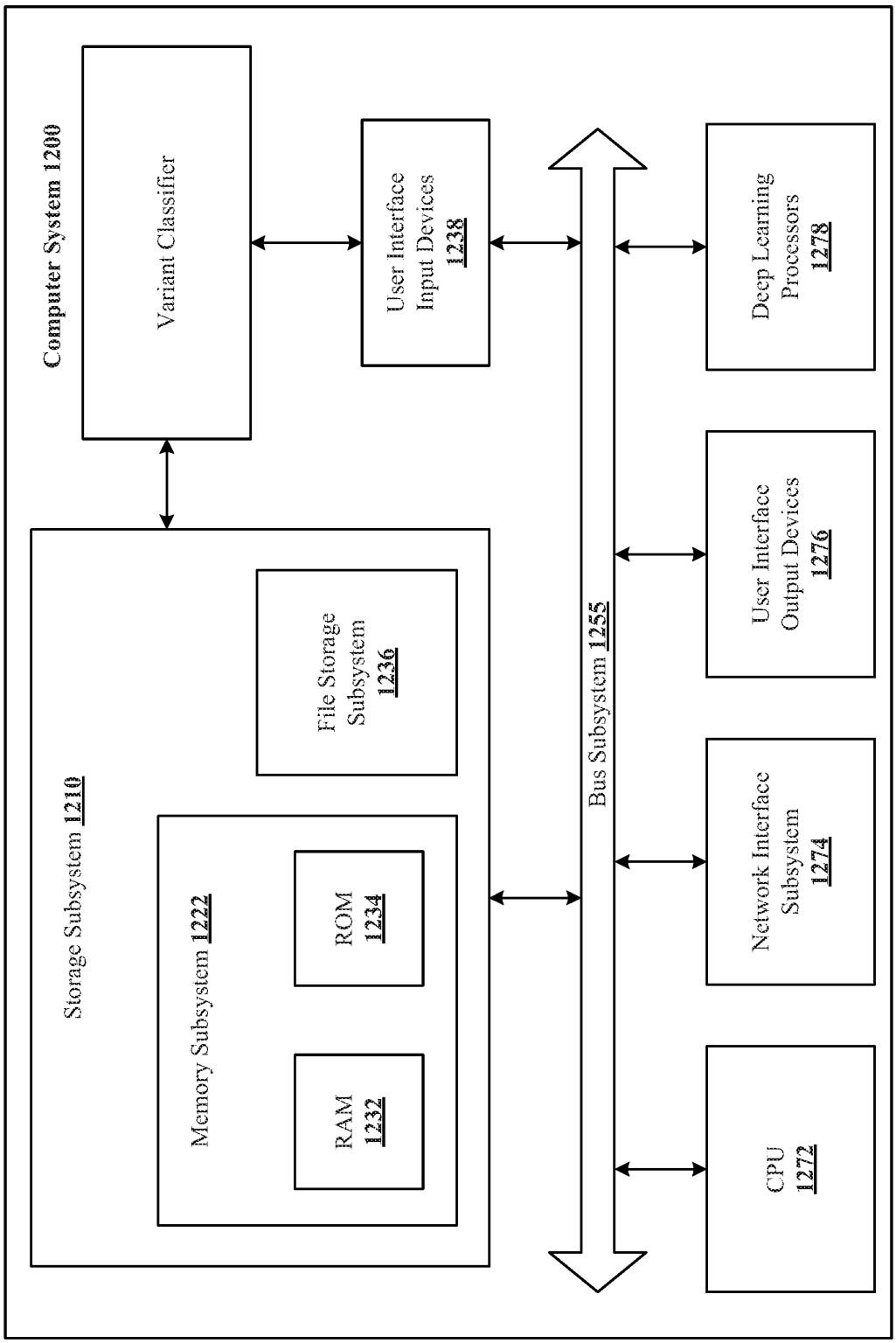
Computer System 1200
Variant Classifier
User Interface Input Devices 1238
Deep Learning Processors 1278
Storage Subsystem 1210
File Storage Subsystem 1236
Memory Subsystem 1222
ROM 1234
RAM 1232
Bus Subsystem 1255
User Interface Output Devices 1276
Network Interface Subsystem 1274
CPU 1272

FIG. 12