



(51) International Patent Classification:

G16B 40/10 (2019.01) G06N 3/02 (2006.01)

G16B 30/00 (2019.01)

(21) International Application Number:

PCT/US2021/036872

(22) International Filing Date:

10 June 2021 (10.06.2021)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

16/899,545 11 June 2020 (11.06.2020) US

(71) Applicant: LIFE TECHNOLOGIES CORPORATION
[US/US]; 5823 Newton Drive, Carlsbad, CA 92008 (US).

(72) Inventors: CHU, Yong; 5823 Newton Drive, Carlsbad, CA 92008 (US). SCHNEIDER, Stephanie, Jo; 5823 Newton Drive, Carlsbad, CA 92008 (US). SCHAEFFER, Ry-

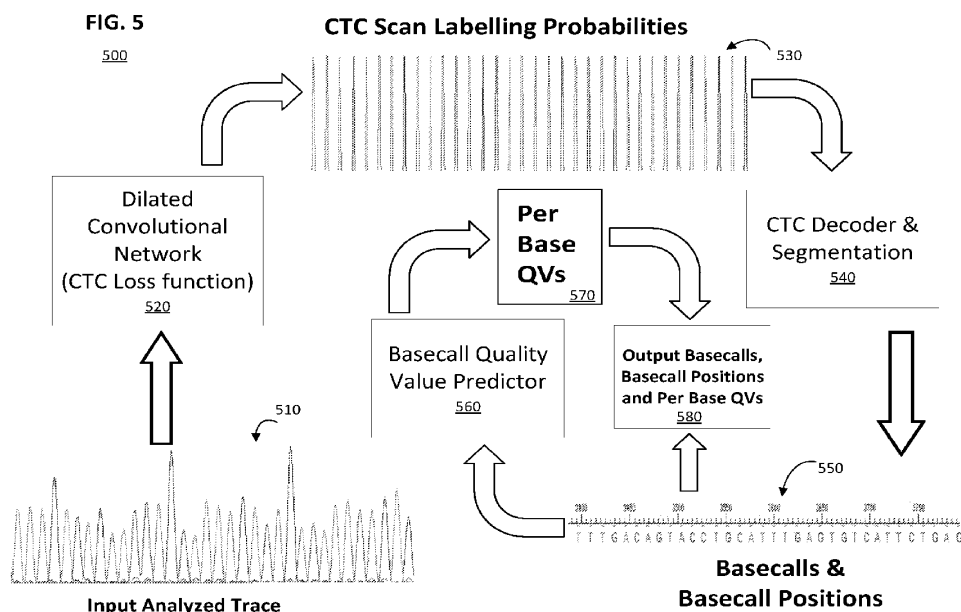
lan; 5823 Newton Drive, Carlsbad, CA 92008 (US). WOO, David; 5823 Newton Drive, Carlsbad, CA 92008 (US).

(74) Agent: LEE, Elaine; Mauriel Kapouytian Woods LLP, 15 W. 26th Street, 7th Floor, New York, NY 10010 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ,

(54) Title: BASECALLER WITH DILATED CONVOLUTIONAL NEURAL NETWORK



(57) Abstract: A method of automatically sequencing or basecalling one or more DNA (deoxyribonucleic acid) molecules of a biological sample is described. The method comprises using a capillary electrophoresis genetic analyzer to measure the biological sample to obtain at least one input trace comprising digital data corresponding to fluorescence values for a plurality of scans. Scan labelling probabilities for the plurality of scans are generated using a trained artificial neural network comprising a plurality of layers including convolutional layers. A basecall sequence comprising a plurality of basecalls for the one or more DNA molecules based on the scan labelling probabilities for the plurality of scans is determined.

UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

- *with international search report (Art. 21(3))*
- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*