

# (12) International Application Status Report

**Received at International Bureau:** 24 October 2019 (24.10.2019)

**Information valid as of:** 26 May 2020 (26.05.2020)

**Report generated on:** 19 September 2020 (19.09.2020)

**(10) Publication number:**

WO2020/115580

**(43) Publication date:**

11 June 2020 (11.06.2020)

**(26) Publication language:**

English (EN)

**(21) Application Number:**

PCT/IB2019/059139

**(22) Filing Date:**

24 October 2019 (24.10.2019)

**(25) Filing language:**

English (EN)

**(31) Priority number(s):**

62/774,494 (US)

**(31) Priority date(s):**

03 December 2018 (03.12.2018)

**(31) Priority status:**

Priority document received (in compliance with PCT Rule 17.1)

**(51) International Patent Classification:**

*G16B 20/30* (2019.01); *G16B 40/20* (2019.01); *G06N 3/04* (2006.01)

**(71) Applicant(s):**

KING ABDULLAH UNIVERSITY OF SCIENCE AND TECHNOLOGY [SA/SA]; 4700 King Abdullah University of Science and Technology Thuwal 23955-6900 (SA) *(for all designated states)*

**(72) Inventor(s):**

GAO, Xin; 4700 King Abdullah University of Science and Technology Thuwal 23955-6900 (SA)

UMAROV, Ramzan; 4700 King Abdullah University of Science and Technology Thuwal 23955-6900 (SA)

**(54) Title (EN):** SYSTEM AND METHOD FOR PROMOTER PREDICTION IN HUMAN GENOME

**(54) Title (FR):** SYSTÈME ET PROCÉDÉ DE PRÉDICTION DE PROMOTEUR DANS LE GÉNOME HUMAIN

**(57) Abstract:**

**(EN):** A method for training a deep neural network model (100) based on a known genome sequence (500) includes receiving (1100) the known genome sequence (500); training (1102) the deep neural network model (100) with a current negative set (502) obtained from the known genome sequence (500); applying (1104) the deep neural network model (100) to the known genome sequence (500) and recording false positive sets; selecting (1106) a subset of the new false positive sets (508); updating (1108) the current negative set (502) with the new false positive sets (508); and repeating (1110) the steps of training, applying, selecting and updating until a number of the new false positive sets is smaller than a given threshold.

**(FR):** La présente invention concerne un procédé d'apprentissage d'un modèle de réseau neuronal profond (100) sur la base d'une séquence génomique connue (500), ledit procédé consistant à recevoir (1100) la séquence génomique connue (500) ; à former (1102) le modèle de réseau neuronal profond (100) avec un ensemble négatif actuel (502) obtenu à partir de la séquence génomique connue (500) ; à appliquer (1104) le modèle de réseau neuronal profond (100) à la séquence génomique connue (500) et à enregistrer de faux ensembles positifs ; à sélectionner (1106) un sous-ensemble des nouveaux ensembles faux positifs (508) ; à mettre à jour (1108) l'ensemble négatif actuel (502) avec les nouveaux ensembles faux positifs (508) ; et à répéter (1110) les étapes d'apprentissage, d'application, de sélection et de mise à jour jusqu'à ce qu'un nombre de nouveaux ensembles faux positifs soit inférieur à un seuil donné.

**International search report:**

Received at International Bureau: 24 January 2020 (24.01.2020) [EP]

**International Report on Patentability (IPRP) Chapter II of the PCT:**

Not available

**(81) Designated States:**

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, 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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW

European Patent Office (EPO) : 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

African Intellectual Property Organization (OAPI) : BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG

African Regional Intellectual Property Organization (ARIPO) : BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW

Eurasian Patent Organization (EAPO) : AM, AZ, BY, KG, KZ, RU, TJ, TM

**Declarations:**

Declaration of inventorship (Rules 4.17(iv) and 51bis.1(a)(iv)) for the purposes of the designation of the United States of America