

【公報種別】特許法第17条の2の規定による補正の掲載

【部門区分】第6部門第3区分

【発行日】令和2年12月10日(2020.12.10)

【公表番号】特表2020-530917(P2020-530917A)

【公表日】令和2年10月29日(2020.10.29)

【年通号数】公開・登録公報2020-044

【出願番号】特願2019-567720(P2019-567720)

【国際特許分類】

G 0 6 N 3/08 (2006.01)

G 0 6 N 20/00 (2019.01)

G 1 6 B 40/20 (2019.01)

【F I】

G 0 6 N 3/08

G 0 6 N 20/00 1 3 0

G 1 6 B 40/20

【手続補正書】

【提出日】令和2年10月14日(2020.10.14)

【手続補正1】

【補正対象書類名】特許請求の範囲

【補正対象項目名】全文

【補正方法】変更

【補正の内容】

【特許請求の範囲】

【請求項1】

病原性一塩基多型(SNP)を特定するコンピュータ実装方法であって、

畳み込みニューラルネットワーク(CNN)のアンサンブルを使用してサイクル内でラベリングされていないSNPを処理するステップであって、前記サイクルは、第1の段階と第2の段階とを含み、

前記ラベリングされていないSNPは、良性とラベリングされておらず、予測されるエリート病原性SNPに付加されていない組合せで生成されたSNPバリエーションを含む、ステップと、

サイクルの前記第1の段階において処理するステップであって、

予測されるエリート病原性SNPがもしあればそれと組み合わせて使用するために、前記ラベリングされていないSNPから前記アンサンブル内でCNN毎に別々の疑似病原性訓練例を選択するステップと、

前記選択された疑似病原性訓練例および前記予測されるエリート病原性SNPがもしあればそれとバランスをとって、SNPの一般的な良性訓練例を使用してCNNの前記アンサンブルを訓練し、それによって前記アンサンブル内の前記CNNのパラメータを更新するステップと

を含む、ステップと、

前記サイクルの前記第2の段階において処理するステップであって、

CNNの前記アンサンブルに前記更新されたパラメータを適用して前記第1の段階に関わっていないラベリングされていないSNPに対する病原性スコアを生成するステップと、

前記病原性スコアを使用して異常値SNPを特定するステップと、

前記異常値SNPを前記予測されるエリート病原性SNP内に累積するステップと

を含む、ステップと、

第1のサイクル内で前記第1の段階および第2の段階の処理を実行するステップであって

予測されるエリート病原性SNPなしで開始するステップと、
前記第2の段階の処理で前記予測されるエリート病原性SNPとして累積された前記特定された異常値SNPで終了するステップと
を含む。ステップと、
第2およびそれ以降のサイクル内で前記第1の段階および第2の段階の処理を繰り返すステップであって、
前のサイクルの前記第2の段階の処理において異常値として特定された前記予測されるエリート病原性SNPで開始するステップと、
前記予測されるエリート病原性SNPで累積された追加の特定された異常値SNPで終了するステップと
を含む。ステップと、
最後のサイクルまで前記サイクルを繰り返すステップであって、その後、前記第2の段階の処理において累積された前記予測されるエリート病原性SNPの数は、前記一般的な良性訓練例におけるヌクレオチドの数の所定の範囲内である、ステップと
を含む。コンピュータ実装方法。

【請求項2】

CNNの前記アンサンブルは各々、深層畳み込みニューラルネットワーク(略してDCNN)である、請求項1に記載のコンピュータ実装方法。

【請求項3】

前記アンサンブルは4から16個のDCNNを含む、請求項2に記載のコンピュータ実装方法。

【請求項4】

前記サイクルにおいてエポック内でDCNNの前記アンサンブルを訓練するステップをさらに含む、請求項2または3に記載のコンピュータ実装方法。

【請求項5】

前記スコアを使用して、DCNNの前記アンサンブルからスコアを合計することによって現在のサイクルの異常値SNPを選択するステップをさらに含む、請求項2から4のいずれか一項に記載のコンピュータ実装方法。

【請求項6】

前記スコアを使用して、DCNNの前記アンサンブルによってスコアリングされた前記SNPの各々に対する最大平均値をとることによって現在のサイクルの異常値SNPを選択するステップをさらに含む、請求項2から5のいずれか一項に記載のコンピュータ実装方法。

【請求項7】

現在のサイクルの間の置換を伴わないサンプリングは、前記現在のサイクルの間の疑似病原性SNPの互いに素の別々の訓練セットをもたらす、請求項1から6のいずれか一項に記載のコンピュータ実装方法。

【請求項8】

前記最後のサイクルは、所定の数のサイクルである、請求項1から7のいずれか一項に記載のコンピュータ実装方法。

【請求項9】

前記パラメータは、少なくとも畳み込みフィルタ重みおよび学習率である、請求項1から8のいずれか一項に記載のコンピュータ実装方法。

【請求項10】

病原性分類器として使用するために前記CNNの前記アンサンブルの中で前記CNNのうちの1つを選択するステップをさらに含む、請求項1から9のいずれか一項に記載のコンピュータ実装方法。

【請求項11】

前記選択されるCNNは、妥当性確認サンプルについて前記アンサンブルの中の他のCNNより予測が優れている、請求項10に記載のコンピュータ実装方法。

【請求項12】

複数の前記サイクルにわたって完成する共通のエリートセットは、少なくとも400000個

の予測されるエリート病原性SNPを有する、請求項1から11のいずれか一項に記載のコンピュータ実装方法。

【請求項13】

各サイクルにおいて、前記予測されるエリート病原性SNPにおける変異率バイアスを防ぐために、良性SNPとサンプリングされた疑似病原性SNPとの間でトリヌクレオチドコンテキストを照合するステップをさらに含む、請求項1から12のいずれか一項に記載のコンピュータ実装方法。

【請求項14】

前記合成セットからの疑似病原性SNPのサンプリングは、各々の連続するサイクルにおいて5%ずつ減少する、請求項1から13のいずれか一項に記載のコンピュータ実装方法。

【請求項15】

メモリに結合される1つまたは複数のプロセッサを備えるシステムであって、前記メモリは病原性一塩基多型(SNP)を特定する請求項1から14のいずれか一項に記載のコンピュータ実装方法に対するコンピュータ命令がロードされているシステム。

【請求項16】

病原性一塩基多型(SNP)を特定するコンピュータ実装方法に対するコンピュータプログラム命令が格納された非一時的コンピュータ可読記憶媒体であって、

前記コンピュータプログラム命令は、プロセッサで実行されたときに、請求項1から14のいずれか一項に記載のコンピュータ実装方法を実行する、非一時的コンピュータ可読記憶媒体。