

(12) DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITÉ DE COOPÉRATION EN MATIÈRE DE BREVETS (PCT)

(19) Organisation Mondiale de la
Propriété Intellectuelle
Bureau international



(43) Date de la publication internationale
23 septembre 2021 (23.09.2021)

WIPO | PCT

(10) Numéro de publication internationale
WO 2021/185826 A1

(51) Classification internationale des brevets :

G05B 23/02 (2006.01) G06N 20/00 (2019.01)
G06F 16/215 (2019.01) G06N 3/08 (2006.01)

LES BORDEAUX 75-77, avenue Marcel Dassault, 33700
CS50108 MERIGNAC (FR).

(21) Numéro de la demande internationale :

PCT/EP2021/056668

(74) Mandataire : **HABASQUE, Etienne** et al. ; LAVOIX, 2,
Place d'Estienne d'Orves, 75441 PARIS CEDEX 09 (FR).

(22) Date de dépôt international :

16 mars 2021 (16.03.2021)

(81) États désignés (sauf indication contraire, pour tout titre de
protection nationale disponible) : 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.

(25) Langue de dépôt :

français

(26) Langue de publication :

français

(30) Données relatives à la priorité :

FR2002546 16 mars 2020 (16.03.2020) FR

(71) Déposant : **THALES** [FR/FR] ; Tour Carpe Diem Place
des Corolles Esplanade Nord, 92400 COURBEVOIE (FR).

(72) Inventeurs : **DUBEDAT, Philippe** ; c/o THALES DMS
France SAS Campus THALES BORDEAUX 75-77, avenue
Marcel Dassault, 33700 CS50108 MERIGNAC (FR). **LEY-
RONAS, Victor** ; c/o THALES DMS France SAS Cam-
pus THALES BORDEAUX 75-77, avenue Marcel Das-
sault, 33700 CS50108 MERIGNAC (FR). **CAILLAUD,
Yannick** ; c/o THALES DMS France SAS Campus THA-

(84) États désignés (sauf indication contraire, pour tout titre de
protection régionale disponible) : ARIPO (BW, GH, GM,
KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG,
ZM, ZW), eurasien (AM, AZ, BY, KG, KZ, RU, TJ, TM),
européen (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

(54) Title: METHOD FOR CONSOLIDATING A DATASET FOR PREDICTIVE MAINTENANCE AND ASSOCIATED DEVICES

(54) Titre : PROCÉDÉ DE CONSOLIDATION D'UN ENSEMBLE DE DONNÉES POUR DE LA MAINTENANCE PRÉDICTIVE
ET DISPOSITIFS ASSOCIÉS

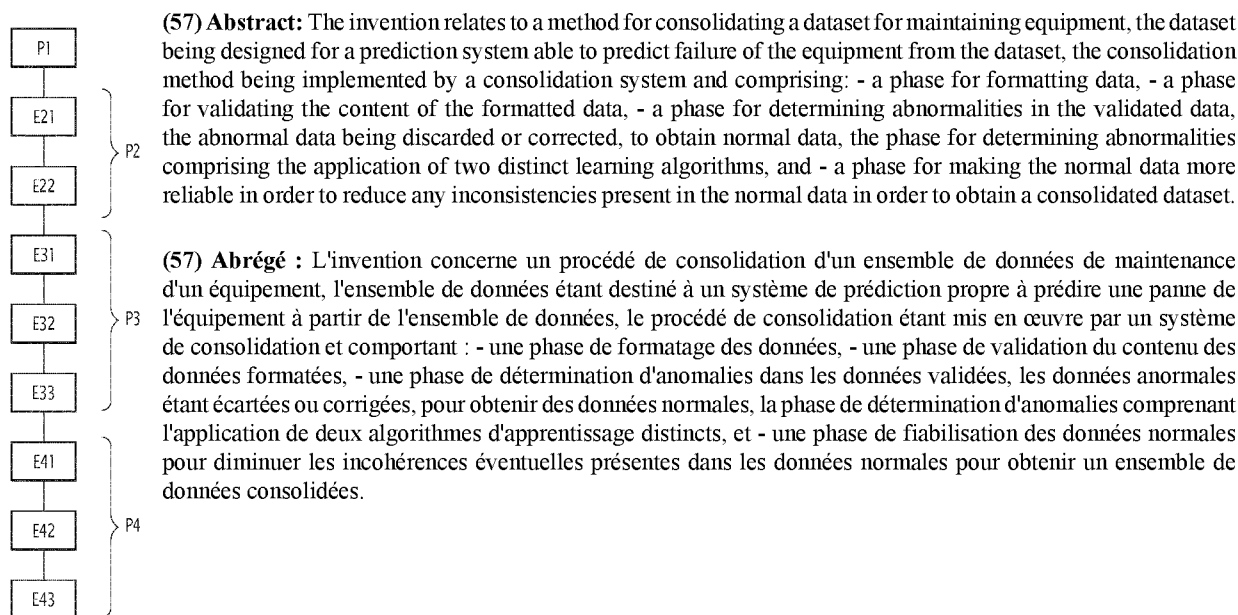


FIG.3

(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML,
MR, NE, SN, TD, TG).

Publiée:

— avec rapport de recherche internationale (Art. 21(3))

Procédé de consolidation d'un ensemble de données pour de la maintenance prédictive et dispositifs associés

5 La présente invention concerne un procédé de consolidation, notamment de la qualité, d'un ensemble de données de maintenance d'un équipement, l'ensemble de données étant destiné à un système de prédiction propre à prédire une panne de l'équipement à partir de l'ensemble de données. La présente invention se rapporte également à un système de prédiction, un produit programme d'ordinateur et un support
10 lisible d'informations associés.

Dans le domaine de la maintenance prédictive, il s'agit d'anticiper un événement rare, tel qu'une panne ou une dépose d'un équipement, avant qu'un tel événement ne se produise. Par définition, un événement rare est un événement survenant, par exemple, moins de 20% du temps de fonctionnement de l'équipement et une dépose de
15 l'équipement est une maintenance ou un changement de l'équipement.

Le rôle d'une telle maintenance prédictive est crucial pour les équipements avioniques dans la mesure où une faille d'un équipement entraîne une mise au sol de l'avion affecté.

L'utilisation d'un système de prédiction permet une telle anticipation, à partir de
20 données de maintenance collectées, par exemple, lors de l'utilisation de l'équipement. Un tel système de prédiction fournit en sortie des données de prédiction représentatives de la probabilité que l'équipement subisse une panne.

Toutefois, en pratique, les systèmes de prédiction connus ne donnent pas entière satisfaction notamment du fait d'un taux de fausse alerte trop élevé, ce qui génère des
25 opérations de maintenance coûteuses et inutiles.

Il existe donc un besoin pour un procédé permettant d'obtenir un système de prédiction présentant une meilleure précision, et notamment un taux de fausse alerte plus faible.

A cet effet, il est proposé un procédé de consolidation, notamment de la qualité,
30 d'un ensemble de données de maintenance d'un équipement, l'ensemble de données étant destiné à un système de prédiction propre à prédire une panne de l'équipement à partir de l'ensemble de données, le procédé de consolidation étant mis en œuvre par un système de consolidation, le procédé comportant une phase de formatage des données, pour obtenir des données formatées, une phase de validation du contenu des données formatées, pour obtenir des données validées, une phase de détermination d'anomalies
35 dans les données validées, les données anormales étant écartées ou corrigées, pour

obtenir des données normales, la phase de détermination d'anomalies comprenant l'application de deux algorithmes d'apprentissage, en particulier d'apprentissage automatique non supervisé, distincts, et une phase de fiabilisation des données normales pour diminuer les incohérences éventuelles présentes dans les données normales pour
5 obtenir un ensemble de données consolidées.

Un tel procédé résulte des travaux originaux du demandeur.

En effet, le demandeur lorsqu'il a été confronté au problème d'améliorer la prédiction de panne a fait face à de multiples voies d'amélioration.

Notamment, le demandeur a beaucoup exploré d'améliorer les calculs menés par
10 le procédé.

Toutefois, lors de cette analyse, le demandeur a identifié un problème plus spécifique.

En effet, il est crucial de disposer de données de maintenance fiables, avant leur exploitation via des systèmes de prédiction.

Cependant, les données collectées dans l'industrie présentent fréquemment des anomalies, telles que des manques, des incohérences ou des champs invalides. L'utilisation de données anormales peut engendrer des biais et fausser les prédictions issues des systèmes de prédiction.
15

Pour lever ces risques, le présent procédé propose un ensemble de modules de traitement, faisant appel à des techniques de machine learning (qui renvoie à la dénomination française d'« apprentissage automatique») non supervisé, pour automatiquement exclure les données anormales, expliquer leurs anomalies, et corriger de manière méthodique les différents problèmes de qualité de la donnée.
20

Suivant des modes de réalisation particuliers, le procédé de consolidation comprend une ou plusieurs des caractéristiques suivantes, prise(s) isolément ou suivant toutes les combinaisons techniquement possibles :
25

- le premier algorithme d'apprentissage est un algorithme de détection d'anomalies et le deuxième algorithme d'apprentissage est un algorithme de localisation d'anomalies.

- l'algorithme de détection d'anomalies est un algorithme de LOF.

- l'algorithme est un filtre de Hampel.

- lors de la phase de détermination d'anomalies, il est mis en œuvre une extraction de caractéristiques à partir des données validées avant l'application des deux algorithmes d'apprentissage.

30

- les caractéristiques extraites sont des concaténations de données ou des données obtenues par application d'opérations mathématiques, en particulier une moyenne arithmétique.

- la phase de fiabilisation des données comporte, pour chaque donnée manquante identifiée dans les données, une étape d'extrapolation des données normales manquantes à partir d'autres données.

- la phase de fiabilisation des données des étapes d'élimination des éventuels doublons et des éventuelles incohérences.

- la phase de validation comporte le test de la conformité des données avec des règles prédéfinies, les données non conformes étant écartées.

La présente description concerne aussi un système de consolidation, notamment de la qualité, d'un ensemble de données de maintenance d'un équipement, l'ensemble de données étant destiné à un système de prédiction propre à prédire une panne de l'équipement à partir de l'ensemble de données, le système de consolidation comprenant une unité de formatage des données propre à formater des données incidentes pour obtenir des données formatées, une unité de validation du contenu des données formatées, pour obtenir des données validées, une unité de détermination d'anomalies dans les données validées, les données anormales étant écartées, pour obtenir des données normales, l'unité de détermination d'anomalies étant propre à appliquer deux algorithmes d'apprentissage, en particulier d'apprentissage automatique non supervisé, distincts, et une unité de fiabilisation des données normales pour diminuer les incohérences éventuelles présentes dans les données normales pour obtenir un ensemble de données consolidées.

La présente description se rapporte aussi à un produit programme d'ordinateur comportant un support lisible d'informations, sur lequel est mémorisé un programme d'ordinateur comprenant des instructions de programme, le programme d'ordinateur étant chargeable sur une unité de traitement de données et adapté pour entraîner la mise en œuvre d'étapes d'un procédé tel que précédemment décrit lorsque le programme d'ordinateur est mis en œuvre sur l'unité de traitement des données.

La présente description concerne aussi un support lisible d'informations comportant des instructions de programme formant un programme d'ordinateur, le programme d'ordinateur étant chargeable sur une unité de traitement de données et adapté pour entraîner la mise en œuvre d'étapes d'un procédé tel que précédemment décrit lorsque le programme d'ordinateur est mis en œuvre sur l'unité de traitement de données.

D'autres caractéristiques et avantages de l'invention apparaîtront à la lecture de la description qui suit de modes de réalisation de l'invention, donnée à titre d'exemple uniquement et en référence aux dessins qui sont :

- figure 1, une vue schématique d'un équipement et d'un système de consolidation d'un point de vue structurel,
- figure 2, une autre représentation schématique de l'équipement et du système de consolidation de la figure 1 d'un point de vue fonctionnel, et
- figure 3, un ordinogramme d'un exemple de mise en œuvre d'un procédé de consolidation.

Un équipement 6 et un système de consolidation 10 sont représentés sur la figure 1.

L'équipement 6 est, par exemple, un équipement électronique tel qu'un équipement avionique.

Alternativement, l'équipement 6 est n'importe quel type d'équipement, électronique ou non, dont des déposes sont à anticiper au cours du temps.

Par exemple, l'équipement 6 est un système de pilotage d'un aéronef. Par définition, un aéronef est un moyen de transport capable de s'élever et de se mouvoir en altitude, au sein de l'atmosphère terrestre. Un hélicoptère ou un avion est un exemple d'aéronef.

Selon l'exemple décrit, l'aéronef est un avion.

En variante, l'équipement 6 fait partie d'un autre système, notamment un système de transport comme un train ou une automobile.

Le système de consolidation 10 est propre à consolider un ensemble de données de maintenance d'un équipement 6 (et en particulier la qualité), l'ensemble de données étant destiné à un système de prédiction propre à prédire une panne de l'équipement 6 à partir de l'ensemble de données

D'un point de vue strictement matériel, le système de consolidation 10 peut être vu comme un calculateur interagissant avec un produit programme d'ordinateur 12 comme schématiquement représenté sur la figure 1.

L'interaction entre le système de consolidation 10 et le produit programme d'ordinateur 12 permet la mise en œuvre d'un procédé de consolidation d'un ensemble de données. Le procédé de consolidation est ainsi un procédé mis en œuvre par ordinateur.

Le système de consolidation 10 est un ordinateur de bureau. En variante, le système de consolidation 10 est un ordinateur monté sur un rack, un ordinateur portable, une tablette, un assistant numérique personnel (PDA) ou un smartphone.

Dans des modes de réalisation spécifiques, l'ordinateur est adapté pour fonctionner en temps réel et/ou est dans un système embarqué, notamment dans un véhicule tel qu'un avion.

Dans le cas de la figure 1, le système de consolidation 10 comprend une unité de calcul 14, une interface utilisateur 16 et un dispositif de communication 18.

L'unité de calcul 14 est un circuit électronique conçu pour manipuler et/ou transformer des données représentées par des quantités électroniques ou physiques dans des registres du système de consolidation 10 et/ou des mémoires en d'autres données similaires correspondant à des données physiques dans les mémoires de registres ou d'autres types de dispositifs d'affichage, de dispositifs de transmission ou de dispositifs de mémorisation.

En tant qu'exemples spécifiques, l'unité de calcul 14 comprend un processeur monocœur ou multicœurs (tel qu'une unité de traitement centrale (CPU), une unité de traitement graphique (GPU), un microcontrôleur et un processeur de signal numérique (DSP)), un circuit logique programmable (comme un circuit intégré spécifique à une application (ASIC), un réseau de portes programmables *in situ* (FPGA), un dispositif logique programmable (PLD) et des réseaux logiques programmables (PLA)), une machine à états, une porte logique et des composants matériels discrets.

L'unité de calcul 14 comprend une unité de traitement de données 20 adaptée pour traiter des données, notamment en effectuant des calculs, des mémoires 22 adaptées à stocker des données et un lecteur 24 adapté à lire un support lisible par ordinateur.

L'interface utilisateur 16 comprend un dispositif d'entrée 26 et un dispositif de sortie 28.

Le dispositif d'entrée 26 est un dispositif permettant à l'utilisateur du système de consolidation 10 de saisir sur le système de consolidation 10 des informations ou des commandes.

Sur la figure 1, le dispositif d'entrée 26 est un clavier. En variante, le dispositif d'entrée 26 est un périphérique de pointage (tel qu'une souris, un pavé tactile et une tablette graphique), un dispositif de reconnaissance vocale, un oculomètre ou un dispositif haptique (analyse des mouvements).

Le dispositif de sortie 28 est une interface utilisateur graphique, c'est-à-dire une unité d'affichage conçue pour fournir des informations à l'utilisateur du système de consolidation 10.

Sur la figure 1, le dispositif de sortie 28 est un écran d'affichage permettant une présentation visuelle de la sortie. Dans d'autres modes de réalisation, le dispositif de

sortie est une imprimante, une unité d'affichage augmenté et/ou virtuel, un haut-parleur ou un autre dispositif générateur de son pour présenter la sortie sous forme sonore, une unité produisant des vibrations et/ou des odeurs ou une unité adaptée à produire un signal électrique.

5 Dans un mode de réalisation spécifique, le dispositif d'entrée 26 et le dispositif de sortie 28 sont le même composant formant des interfaces homme-machine, tel qu'un écran interactif.

Le dispositif de communication 18 permet une communication unidirectionnelle ou bidirectionnelle entre les composants du système de consolidation 10. Par exemple, le
10 dispositif de communication 18 est un système de communication par bus ou une interface d'entrée / sortie.

La présence du dispositif de communication 18 permet que, dans certains modes de réalisation, les composants du système de consolidation 10 soient distants les uns des autres.

15 Le produit programme informatique 12 comprend un support lisible par ordinateur 32.

Le support lisible par ordinateur 32 est un dispositif tangible lisible par le lecteur 24 de la calculatrice 14.

Notamment, le support lisible par ordinateur 32 n'est pas un signal transitoire en soi, tels que des ondes radio ou d'autres ondes électromagnétiques à propagation libre, telles que des impulsions lumineuses ou des signaux électroniques.

Un tel support de stockage lisible par ordinateur 32 est, par exemple, un dispositif de stockage électronique, un dispositif de stockage magnétique, un dispositif de stockage optique, un dispositif de stockage électromagnétique, un dispositif de stockage à semi-
25 conducteur ou toute combinaison de ceux-ci.

En tant que liste non exhaustive d'exemples plus spécifiques, le support de stockage lisible par ordinateur 32 est un dispositif codé mécaniquement, tel que des cartes perforées ou des structures en relief dans une gorge, une disquette, un disque dur, une mémoire morte (ROM), une mémoire vive (RAM), une mémoire effaçable
30 programmable en lecture seule (EPROM), une mémoire effaçable électriquement et lisible (EEPROM), un disque magnéto-optique, une mémoire vive statique (SRAM), un disque compact (CD-ROM), un disque numérique polyvalent (DVD), une clé USB, un disque souple, une mémoire flash, un disque à semi-conducteur (SSD) ou une carte PC telle qu'une carte mémoire PCMCIA.

Un programme d'ordinateur est stocké sur le support de stockage lisible par ordinateur 32. Le programme d'ordinateur comprend une ou plusieurs séquences d'instructions de programme mémorisées.

De telles instructions de programme, lorsqu'elles sont exécutées par l'unité de traitement de données 20, entraînent l'exécution d'étapes du procédé d'analyse.

Par exemple, la forme des instructions de programme est une forme de code source, une forme exécutable par ordinateur ou toute forme intermédiaire entre un code source et une forme exécutable par ordinateur, telle que la forme résultant de la conversion du code source via un interpréteur, un assembleur, un compilateur, un éditeur de liens ou un localisateur. En variante, les instructions de programme sont un microcode, des instructions firmware, des données de définition d'état, des données de configuration pour circuit intégré (par exemple du VHDL) ou un code objet.

Les instructions de programme sont écrites dans n'importe quelle combinaison d'un ou de plusieurs langages, par exemple un langage de programmation orienté objet (C++, JAVA, Python), un langage de programmation procédural (langage C par exemple).

Alternativement, les instructions du programme sont téléchargées depuis une source externe via un réseau, comme c'est notamment le cas pour les applications. Dans ce cas, le produit programme d'ordinateur comprend un support de données lisible par ordinateur sur lequel sont stockées les instructions de programme ou un signal de support de données sur lequel sont codées les instructions de programme.

Dans chaque cas, le produit programme d'ordinateur 12 comprend des instructions qui peuvent être chargées dans l'unité de traitement de données 20 et adaptées pour provoquer l'exécution du procédé d'analyse lorsqu'elles sont exécutées par l'unité de traitement de données 20. Selon les modes de réalisation, l'exécution est entièrement ou partiellement réalisée soit sur le système de consolidation 10, c'est-à-dire un ordinateur unique, soit dans un système distribué entre plusieurs ordinateurs (notamment via l'utilisation de l'informatique en nuage).

D'un point de vue fonctionnel, ainsi que visible à la figure 2, le système de consolidation 10 en fonctionnement peut s'interpréter comme comportant une unité de formatage 40, une unité de validation 42, une unité de détermination d'anomalies 44 et une unité de fiabilisation 46.

L'unité de validation 42 comporte un module de normalisation 48, un module d'évaluation 50.

L'unité de détermination d'anomalies 44 comporte un module d'extraction 52, un module de détection d'anomalies 54 et un module de localisation d'anomalies 56.

L'unité de fiabilisation 46 comporte un module d'extrapolation 58, un premier module d'élimination 60 et un deuxième module d'élimination 62.

Le fonctionnement du système de consolidation 10 sera maintenant mieux compris en référence à la figure 3 qui est un ordinogramme illustrant un exemple de mise en œuvre du procédé de consolidation.

Le procédé de consolidation comporte une phase de formatage P1, une phase de validation P2, une phase de détermination d'anomalies P3 et une phase de fiabilisation P4.

Le procédé de consolidation s'applique à des données brutes pouvant provenir de plusieurs endroits, notamment un atelier de réparation ou d'éléments survenus en fonctionnement de l'équipement 6, plus particulièrement lors d'un vol.

Lors de la phase de formatage P1, l'unité de formatage 40 formate des données incidentes.

L'unité de formatage 40 prend en entrée des données brutes collectées et donne en sortie des données formatées selon un format prédéfini.

Selon l'exemple décrit, le format prédéfini est un tableau présentant des colonnes avec des types de données prédéfinis. Par exemple, chaque entrée d'une colonne est un nombre ou chaque entrée d'une colonne est un texte.

Dans la suite de la description, un tableau correspondant au format prédéfini qui rassemble plusieurs données est appelé un jeu de données.

L'unité de formatage 40 assure qu'une donnée brute soit convertie au format prédéfini sans modification du fond de la donnée.

L'unité de formatage 40 applique ainsi aux données brutes un algorithme d'extraction d'information et de restructuration selon le format prédéfini.

Un exemple de tel algorithme est un algorithme qui transforme des données au format xml en tableau. S'il est implémenté dans le langage Python, il peut faire appel au package lxml afin d'analyser les xml. L'unité de formatage 40 convertit également l'encodage des données pour conformer l'encodage à l'encodage prédéfini.

Lorsque l'unité de formatage 40 est dans l'incapacité de formater les données brutes, l'unité de formatage 40 met l'ensemble des données brutes destinés à former un jeu en quarantaine dans l'attente d'une intervention manuelle pour déterminer pourquoi les données brutes ne peuvent pas se conformer au format prédéfini.

La phase de validation P2 comporte une étape de normalisation E21 et une étape d'évaluation E22.

Lors de la phase de validation P2, l'unité de validation 42 filtre les données formatées pour ne conserver que des données valides.

Dans le contexte du présent procédé, une donnée est valide lorsque le contenu de la donnée est conforme à un ensemble de règles prédéfinies.

Les règles prédéfinies sont le plus souvent des règles syntaxiques.

Plusieurs exemples de règles sont donnés dans ce qui suit.

5 Selon un premier exemple, une colonne intitulée « nom du pays » impose comme règle prédéfinie que tous les noms de pays présents dans la colonne appartiennent à une liste de pays connus.

Un deuxième exemple de règle prédéfinie est qu'une colonne soit remplie uniquement de nombre entier compris entre deux valeurs. A titre d'illustration, c'est le cas
10 pour une colonne qui donne l'âge de techniciens.

Selon un troisième exemple, il est exigé que tous les champs contenant des poids soient exprimés en kilogrammes.

L'étape de normalisation E21 vise à respecter les règles prédéfinies précitées.

Le module de normalisation 48 convertit les données pour respecter la ou les
15 règles prédéfinies.

En ce sens, le module de normalisation 48 est un module de standardisation.

A titre d'exemple, le module de normalisation 48 est propre à remplacer une donnée par une donnée synonyme ou à convertir une donnée fournie en une première unité en une deuxième unité.

20 Le module de normalisation 48 permet ainsi d'obtenir des jeux de données qui vérifient les règles prédéfinies.

Lors de l'étape d'évaluation E22, le module d'évaluation 50 teste si chaque jeu de données vérifie les règles prédéfinies.

Si une seule des règles prédéfinies n'est pas respectée, le module d'évaluation 50
25 envoie le jeu de données en quarantaine en attente d'une intervention manuelle.

Si toutes les règles prédéfinies sont respectées, le module d'évaluation 50 valide le jeu de données.

A l'issue de la phase de validation P2, il est ainsi obtenu des jeux de données valides, chaque jeu de données invalide ayant été écarté.

30 La phase de détermination d'anomalies P3 comporte une étape d'extraction E31, une étape de détection d'anomalies E32, une étape de localisation d'anomalies E33.

Lors de l'étape d'extraction E31, le module d'extraction 52 associe à chaque jeu de données un ensemble de caractéristiques.

Par exemple, pour des données qui sont des variables numériques, les
35 caractéristiques peuvent être la valeur moyenne (arithmétique notamment) des données, la valeur minimale des données et la valeur maximale des données.

Un exemple de telles données numériques est l'âge du technicien en charge de la réparation ou la durée de son expérience de la profession.

Selon un autre exemple, certaines caractéristiques peuvent être calculées en utilisant plusieurs variables. À titre d'illustration, dans le cas où le jeu de données comporte deux variables, une caractéristique peut être le nombre de lignes où les deux variables diffèrent.

A l'issue de l'étape d'extraction E31, il est ainsi obtenu un ensemble de caractéristiques extraites.

Lors de l'étape de détection d'anomalies E32, le module de détection d'anomalies 54, simplement noté module de détection 54 dans ce qui suit, est propre à détecter une anomalie dans un jeu de données.

Plus précisément, le module de détection 54 est propre à appliquer un algorithme sur des données incidentes organisées en jeux de données.

L'algorithme appliqué par le module de détection 54 est un algorithme de détection d'anomalies. Un tel algorithme est souvent désigné sous la dénomination anglaise correspondante de « *outlier detection* ».

Un tel algorithme permet de déterminer les données qui s'écartent de l'habitude.

L'algorithme du module de détection 54 est un algorithme d'apprentissage non supervisée, notamment un algorithme d'apprentissage automatique non supervisée.

L'algorithme prend en entrée les caractéristiques obtenues à l'étape d'extraction pour un jeu de données à évaluer et les caractéristiques obtenues à l'étape d'extraction pour des jeux de données antérieurs au jeu de données à évaluer et donne en sortie un score.

Le fait de disposer des caractéristiques obtenues à l'étape d'extraction pour des jeux de données antérieurs permet à l'algorithme de constituer des valeurs habituelles pour chaque caractéristique.

Le score indique la déviation des caractéristiques du jeu de données à examiner par rapport aux valeurs habituelles.

Autrement formulé, le score est une quantification de l'anormalité du jeu de données à évaluer.

Par exemple, le score est un nombre entre 0 et 1.

L'algorithme du module de détection 54 apprend sur une base de données historique et est appliqué sur les jeux de données à évaluer.

Plus précisément, pour tenir compte de l'évolution des jeux de données dans le temps, l'algorithme de détections d'anomalie réapprend à chaque nouveau jeu de données incidents.

Un tel algorithme est bien adapté pour détecter des anomalies subtiles mais ne fournit aucune indication sur la nature de l'anomalie.

Un exemple particulier de tel algorithme de détection d'anomalies est l'algorithme de LOF. L'acronyme LOF renvoie à la dénomination anglaise de « Local Outlier Factor »
5 qui signifie littéralement facteur local d'anomalies.

L'algorithme de LOF utilise avantageusement une mesure de la déviation locale de densité d'un jeu de données par rapport aux anciens jeux de données.

Pour illustrer le fonctionnement du module de détection 54 sur un exemple
particulier, il peut être utilisé l'exemple précédent concernant l'âge des techniciens qui
10 serait une donnée numérique dont la caractéristique « valeur moyenne » est mise en
entrée de l'algorithme du module de détection 54.

L'algorithme compare alors avec les moyennes des âges provenant des
précédents jeux de données, l'algorithme vérifie si le jeu de données à évaluer est
similaire ou différente.

Notamment, l'algorithme examine si la moyenne des âges des techniciens dans le
15 jeu de données à évaluer a augmenté significativement par rapport aux moyennes des
précédents jeux de données. Similairement, l'algorithme examine si la moyenne a
diminué significativement.

Si le jeu de données à évaluer est détecté comme anormal, la sortie de
20 l'algorithme correspondra à un score indiquant que le jeu de données à évaluer est
anormal.

D'autres algorithmes peuvent être utilisés ici comme un réseau de neurones ou
une machine à support de vecteur.

A l'issue de l'étape de détection d'anomalies E32, pour chaque jeu de données, il
25 a été déterminé la présence ou non d'une anomalie.

Lors de l'étape de recherche d'anomalies E33, pour chaque jeu de données
détecté comme anormal, le module de localisation d'anomalies 56, dit simplement module
de localisation 56 dans ce qui suit, cherche à déterminer la ou les colonne(s) anormale(s)
du jeu de données.

En d'autres termes, le module de localisation 56 complète le module de détection
30 54 en fournissant une localisation précise de l'anomalie, et par là une explication au fait
que le jeu de données soit considéré comme anormal.

Le module de localisation 56 applique un algorithme sur chaque caractéristique
d'un jeu de données anormal indépendamment pour déterminer si la caractéristique est
35 responsable de l'anomalie détectée par le module de détection 54.

Un exemple de tel algorithme est un algorithme de filtre de Hampel. Un tel algorithme est plus souvent désigné sous la dénomination anglaise correspondante de « Hampel Filter ».

L'algorithme de filtre de Hampel fonctionne de la manière suivante : pour une caractéristique donnée, une fenêtre temporelle est fixée par l'algorithme (6 mois par exemple) et une médiane glissante est effectuée par l'algorithme en se basant sur les extractions de caractéristiques précédentes. L'algorithme calcule de même un écart médian qui mesure à quel point la caractéristique varie habituellement dans les jeux de données précédents. La valeur de la caractéristique dans la nouvelle extraction est alors comparée à la valeur médiane actuelle : si l'écart dépasse l'écart médian une alerte est lancée pour la colonne concernée.

Un tel algorithme permet ainsi de détecter automatiquement les colonnes responsables des anomalies détectées avec le module de détection 54.

A l'issue de l'étape de recherche E33, la colonne responsable (les colonnes dans certains cas) de l'anomalie de chaque jeu de données anormal est trouvée.

A l'issue de la phase de détermination d'anomalies P3, il reste uniquement les données normales qui sont envoyées à la phase de fiabilisation P4.

La phase de fiabilisation P4 comporte une étape d'extrapolation E41, une première étape d'élimination E42 et une deuxième étape d'élimination E43.

Lors de la phase de fiabilisation P4, l'unité de fiabilisation vise à transformer les données normales en données de confiance, les données de confiance pouvant être considérées comme des données de référence.

Lors de l'étape d'extrapolation E41, le module d'extrapolation 58 complète les données manquantes du jeu de données.

Pour cela, le module d'extrapolation 58 détecte une donnée manquante et extrapole les autres données pour obtenir une donnée manquante.

Les autres données sont, par exemple, les mêmes données de jeux plus anciens ou des données de même nature du jeu de données.

A titre d'illustration, l'extrapolation est mise en œuvre par application d'une moyenne, notamment une moyenne arithmétique.

Lors de la première étape d'élimination, le module d'élimination 60 élimine les données contradictoires.

Pour cela, le premier module d'élimination 60 détecte les données contradictoires à l'aide d'informations connues.

Un exemple de données contradictoire est suivant : dans un jeu de données, une colonne indique qu'un technicien a 20 ans à une date donnée et une autre colonne

indique qu'à la même date, le même technicien a 30 ans d'ancienneté dans son poste. Puis, le premier module d'élimination 60 résout la contradiction, par exemple en éliminant une des deux données.

Dans l'exemple proposé, pour lever la contradiction détectée, il est utilisé une donnée complémentaire qui est la date de l'embauche dudit technicien.

Ensuite, le deuxième module d'élimination 62 met en œuvre la deuxième étape d'élimination E43.

Le deuxième module d'élimination 62 cherche à détecter les doublons et à éliminer l'occurrence inutile.

A l'issue de la quatrième phase de fiabilisation P4, les données sont des données de confiance.

Le procédé de consolidation garantit une fiabilisation fortement accrue des données mémorisées. Les données, ou plus précisément les jeux de données, peuvent alors être utilisées de manière sereine, en limitant les risques de fausser les analyses ultérieures.

Plus précisément, le procédé de consolidation garantit la validité des données (c'est-à-dire la conformité à un schéma de données précis et des règles syntaxiques prédéfinies), la complétude (les données sont dites complètes s'il n'y a pas de manque d'information, tels que des champs vides ou des lignes préfiltrées), l'unicité des données, l'absence d'incohérence dans l'ensemble des données, la précision des données et la non-obsolésence des données.

Le procédé permet ainsi d'entraîner le système de prédiction sur des données de meilleure qualité.

Ceci implique que le procédé permet d'obtenir un système de prédiction présentant une meilleure précision, et notamment un taux de fausse alerte plus faible.

Le procédé peut également être utilisé pour prédire une panne de l'équipement 6 à partir de l'ensemble de données, les données étant alors des données pertinentes pour la prédiction, par exemple des données sélectionnées sur une période de temps prédéfini avant l'instant de prédiction, par exemple 40 vols antérieurs au prochain vol pour lequel on cherche à prédire si un équipement 6 va tomber en panne.

En outre, le procédé est compatible avec tout type de données intervenant dans la maintenance. En effet, les données de maintenance d'un équipement incluent toute donnée ayant un impact sur le fonctionnement de l'équipement 6. En particulier, les données de maintenance comprennent les données de fabrication de l'équipement 6 (notamment où ? quand ? qui ? comment ?), les données en utilisation de l'équipement 6 (notamment le suivi du comportement par des capteurs, par exemple, le comportement

thermique, le temps d'utilisation, la consommation ou autres) et les données d'entretien et/ou de réparation de l'équipement 6 (notamment où ? quand ? qui ? comment ?). Dans certains cas, les données de maintenance seront limitées à l'un des types de données précitées.

- 5 Par exemple, pour le cas de données issues de capteurs, le procédé pourra consister à vérifier que les mesures sont comprises dans un intervalle possible et ne résultent pas d'une erreur d'un capteur ou encore que les mesures sont exprimées dans la même unité, qu'il ne manque pas certaines valeurs (quitte à les interpoler sur la base de la mesure de l'instant antérieur et de l'instant postérieur). Le procédé peut également
- 10 prendre en compte des adaptations d'un capteur à un autre ou encore des dérives temporelles d'un capteur.

REVENDEICATIONS

1. Procédé de consolidation, notamment de la qualité, d'un ensemble de données de maintenance d'un équipement (6), l'ensemble de données étant destiné à un système de prédiction propre à prédire une panne de l'équipement (6) à partir de l'ensemble de données, le procédé de consolidation étant mis en œuvre par un système de consolidation (10), le procédé comportant :

- une phase de formatage des données, pour obtenir des données formatées,
- une phase de validation du contenu des données formatées, pour obtenir des données validées,
- une phase de détermination d'anomalies dans les données validées, les données anormales étant écartées ou corrigées, pour obtenir des données normales, la phase de détermination d'anomalies comprenant l'application de deux algorithmes d'apprentissage, en particulier d'apprentissage automatique non supervisé, distincts, et
- une phase de fiabilisation des données normales pour diminuer les incohérences éventuelles présentes dans les données normales pour obtenir un ensemble de données consolidées.

2. Procédé de consolidation selon la revendication 1, dans lequel le premier algorithme d'apprentissage est un algorithme de détection d'anomalies et le deuxième algorithme d'apprentissage est un algorithme de localisation d'anomalies.

3. Procédé de consolidation selon la revendication 2, dans lequel l'algorithme de détection d'anomalies est un algorithme de LOF.

4. Procédé de consolidation selon la revendication 2 ou 3, dans lequel l'algorithme est un filtre de Hampel.

5. Procédé de consolidation selon l'une quelconque des revendications 1 à 4, dans lequel, lors de la phase de détermination d'anomalies, il est mis en œuvre une extraction de caractéristiques à partir des données validées avant l'application des deux algorithmes d'apprentissage.

6. Procédé de consolidation selon la revendication 5, dans lequel les caractéristiques extraites sont des concaténations de données ou des données obtenues par application d'opérations mathématiques, en particulier une moyenne arithmétique.

5 7. Procédé de consolidation selon l'une quelconque des revendications 1 à 6, dans lequel la phase de fiabilisation des données comporte, pour chaque donnée manquante identifiée dans les données, une étape d'extrapolation des données normales manquantes à partir d'autres données.

10 8. Procédé de consolidation selon l'une quelconque des revendications 1 à 7, dans lequel la phase de fiabilisation des données des étapes d'élimination des éventuels doublons et des éventuelles incohérences.

15 9. Procédé de consolidation selon l'une quelconque des revendications 1 à 8, dans lequel la phase de validation comporte le test de la conformité des données avec des règles prédéfinies, les données non conformes étant écartées.

20 10. Système de consolidation (10), notamment de la qualité, d'un ensemble de données de maintenance d'un équipement (6), l'ensemble de données étant destiné à un système de prédiction propre à prédire une panne de l'équipement (6) à partir de l'ensemble de données, le système de consolidation (10) comprenant:

- une unité de formatage des données propre à formater des données incidentes pour obtenir des données formatées,
- une unité de validation du contenu des données formatées, pour obtenir des données validées,
- une unité de détermination d'anomalies dans les données validées, les données anormales étant écartées, pour obtenir des données normales, l'unité de détermination d'anomalies étant propre à appliquer deux algorithmes d'apprentissage, en particulier d'apprentissage automatique non supervisé, distincts, et
- une unité de fiabilisation des données normales pour diminuer les incohérences éventuelles présentes dans les données normales pour obtenir un ensemble de données consolidées.

11. Produit programme d'ordinateur (12) comportant un programme d'ordinateur comprenant des instructions de programme, le programme d'ordinateur étant chargeable sur une unité de traitement de données (20) et adapté pour entraîner la mise en œuvre d'étapes d'un procédé selon l'une quelconque des revendications 1 à 9 lorsque le programme d'ordinateur est mis en œuvre sur l'unité de traitement des données (20).

12. Support lisible d'informations (32) comportant des instructions de programme formant un programme d'ordinateur, le programme d'ordinateur étant chargeable sur une unité de traitement de données (20) et adapté pour entraîner la mise en œuvre d'étapes d'un procédé selon l'une quelconque des revendications 1 à 9 lorsque le programme d'ordinateur est mis en œuvre sur l'unité de traitement de données (20).

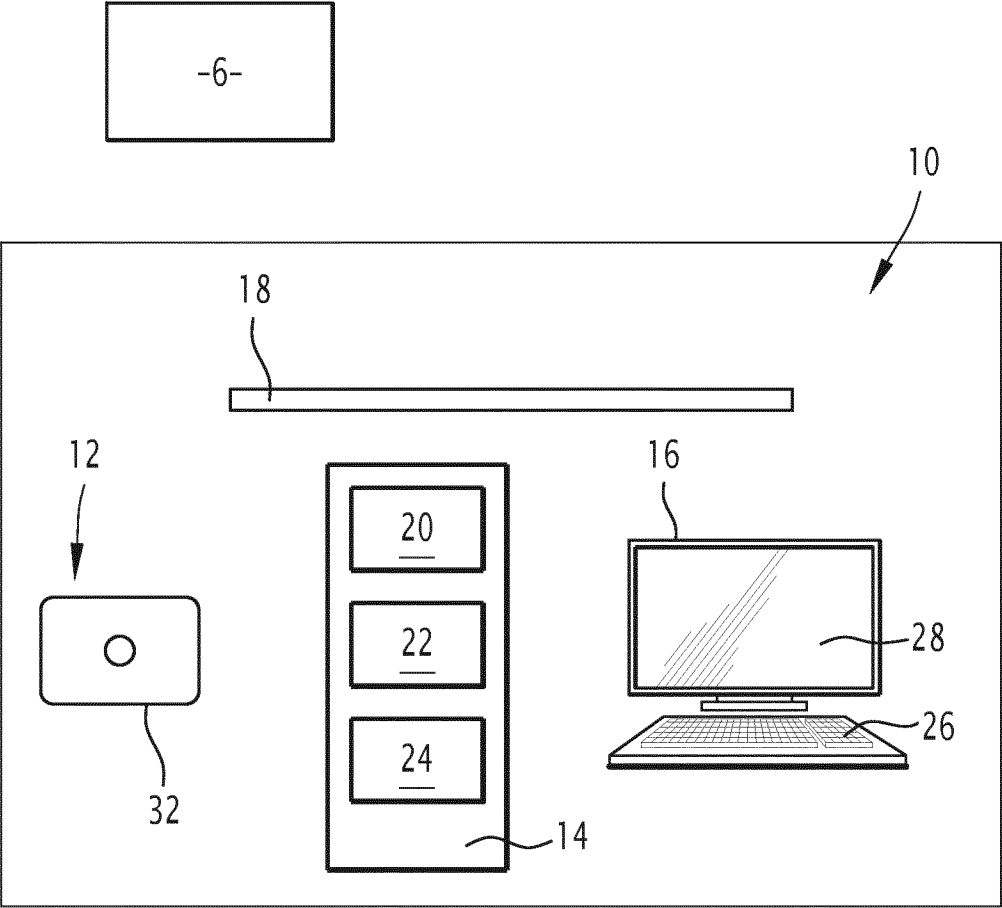


FIG.1

2/3

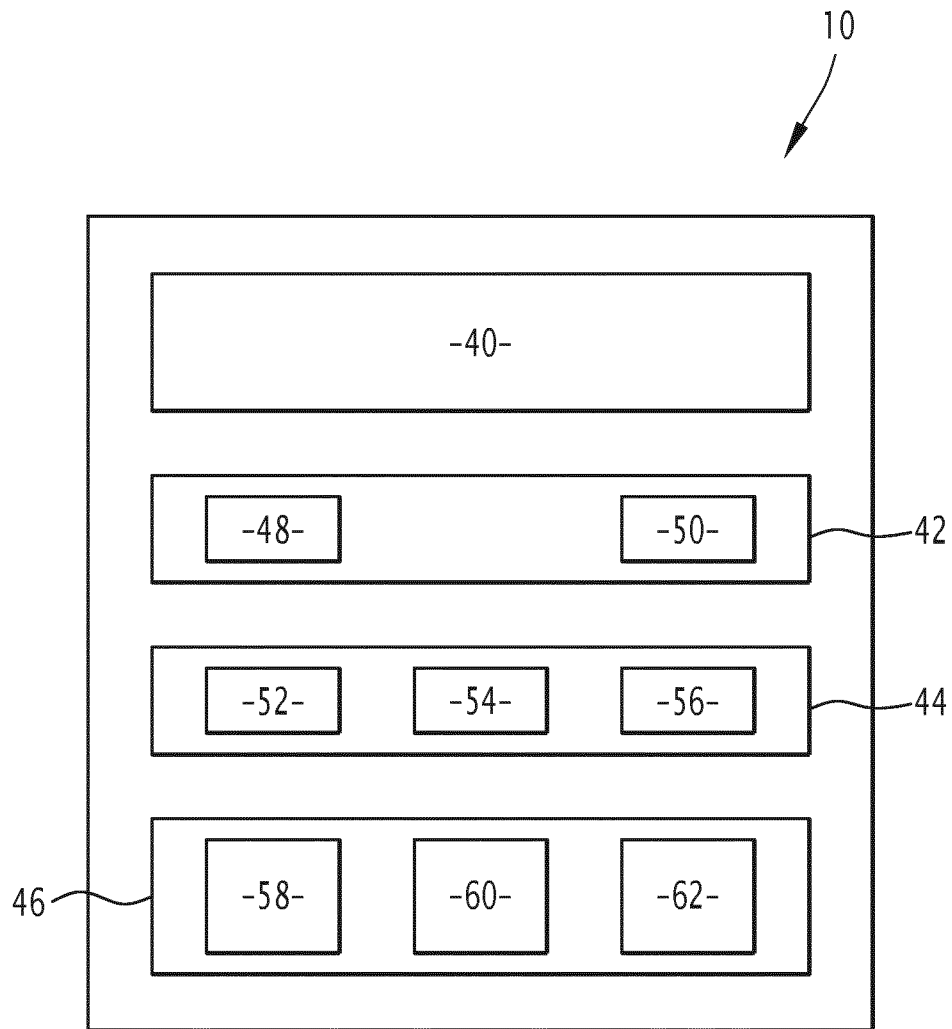
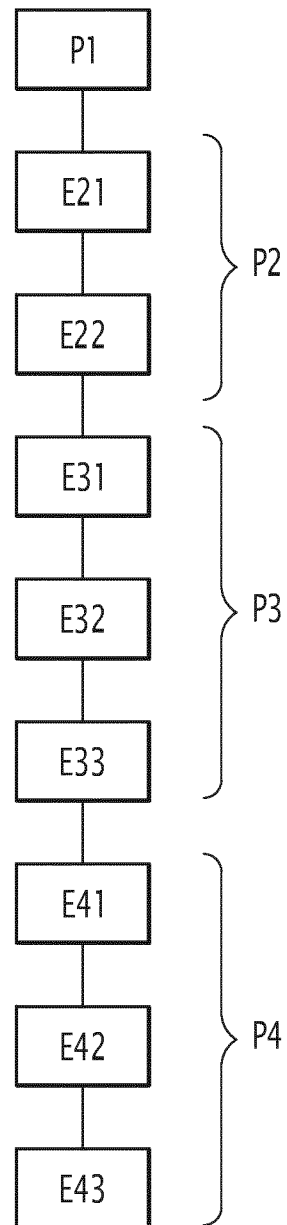


FIG.2

3/3

**FIG.3**

INTERNATIONAL SEARCH REPORT

International application No.

PCT/EP2021/056668

A. CLASSIFICATION OF SUBJECT MATTER**G05B 23/02**(2006.01)i; **G06F 16/215**(2019.01)i; **G06N 20/00**(2019.01)i; **G06N 3/08**(2006.01)n

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G05B; G06F; G06N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CHENG ZHANGYU CHENGZY@WHUT EDU CN ET AL. "Outlier detection using isolation forest and local outlier factor" <i>PROCEEDINGS OF THE CONFERENCE ON RESEARCH IN ADAPTIVE AND CONVERGENT SYSTEMS, ACPUB27, NEW YORK, NY, USA</i> , 24 September 2019 (2019-09-24), pages 161-168 DOI: 10.1145/3338840.3355641 ISBN: 978-1-4503-6843-8. XP058471793 page 161 - page 163	1-12
A	CN 110471946 A (UNIV GUANGDONG TECHNOLOGY) 19 November 2019 (2019-11-19) the whole document	1-3,10-12
A	EP 3107000 A2 (TATA CONSULTANCY SERVICES LTD [IN]) 21 December 2016 (2016-12-21) paragraph [0001] - paragraph [0009] paragraph [0018] - paragraph [0019]	4

☒ Further documents are listed in the continuation of Box C.☒ See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

28 May 2021

Date of mailing of the international search report

08 June 2021

Name and mailing address of the ISA/EP

European Patent Office
p.b. 5818, Patentlaan 2, 2280 HV Rijswijk
Netherlands

Telephone No. (+31-70)340-2040

Facsimile No. (+31-70)340-3016

Authorized officer

Kuntz, Jean-Marc

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/EP2021/056668

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	XU ZEKUN ET AL. "Automatic Hyperparameter Tuning Method for Local Outlier Factor, with Applications to Anomaly Detection" <i>2019 IEEE INTERNATIONAL CONFERENCE ON BIG DATA (BIG DATA)</i> , IEEE, 09 December 2019 (2019-12-09), pages 4201-4207 DOI: 10.1109/BIGDATA47090.2019.9006151 XP033721319 the whole document	1-3,10-12
A	LIGHTFOOT LEONARD ET AL. "The effects of synthetically augmented training data on parameter tuning for anomaly detection algorithms" <i>DEFENSE TRANSFORMATION AND NET-CENTRIC SYSTEMS 2011, SPIE, 1000 20TH ST. BELLINGHAM WA 98225-6705 USA</i> , Vol. 8062, No. 1, 13 May 2011 (2011-05-13), pages 1-11 DOI: 10.1117/12.883526 XP060014074 the whole document	1,10-12
A	ANTUNES ANTONIO LORVAO ET AL. "Adding Value to Sensor Data of Civil Engineering Structures: Automatic Outlier Detection" <i>2018 9TH INTERNATIONAL CONFERENCE ON INFORMATION, INTELLIGENCE, SYSTEMS AND APPLICATIONS (IISA)</i> , IEEE, 23 July 2018 (2018-07-23), pages 1-6 DOI: 10.1109/IISA.2018.8633586 XP033511174 the whole document	1,10-12

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/EP2021/056668

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	110471946	A	19 November 2019	NONE			
EP	3107000	A2	21 December 2016	AU	2016204093	A1	12 January 2017
				CN	106257438	A	28 December 2016
				EP	3107000	A2	21 December 2016
				JP	6611677	B2	27 November 2019
				JP	2017010552	A	12 January 2017
				US	2016371228	A1	22 December 2016

RAPPORT DE RECHERCHE INTERNATIONALE

Demande internationale n°

PCT/EP2021/056668

A. CLASSEMENT DE L'OBJET DE LA DEMANDE INV. G05B23/02 G06F16/215 G06N20/00 ADD. G06N3/08		
Selon la classification internationale des brevets (CIB) ou à la fois selon la classification nationale et la CIB		
B. DOMAINES SUR LESQUELS LA RECHERCHE A PORTE Documentation minimale consultée (système de classification suivi des symboles de classement) G05B G06F G06N		
Documentation consultée autre que la documentation minimale dans la mesure où ces documents relèvent des domaines sur lesquels a porté la recherche		
Base de données électronique consultée au cours de la recherche internationale (nom de la base de données, et si cela est réalisable, termes de recherche utilisés) EPO-Internal, WPI Data		
C. DOCUMENTS CONSIDERES COMME PERTINENTS		
Catégorie*	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
X	CHENG ZHANGYU CHENGZY@WHUT EDU CN ET AL: "Outlier detection using isolation forest and local outlier factor", PROCEEDINGS OF THE CONFERENCE ON RESEARCH IN ADAPTIVE AND CONVERGENT SYSTEMS, ACPUB27, NEW YORK, NY, USA, 24 septembre 2019 (2019-09-24), pages 161-168, XP058471793, DOI: 10.1145/3338840.3355641 ISBN: 978-1-4503-6843-8 page 161 - page 163	1-12
A	----- CN 110 471 946 A (UNIV GUANGDONG TECHNOLOGY) 19 novembre 2019 (2019-11-19) le document en entier ----- -/-	1-3, 10-12
☒ Voir la suite du cadre C pour la fin de la liste des documents ☒ Les documents de familles de brevets sont indiqués en annexe		
* Catégories spéciales de documents cités: "A" document définissant l'état général de la technique, non considéré comme particulièrement pertinent "E" document antérieur, mais publié à la date de dépôt international ou après cette date "L" document pouvant jeter un doute sur une revendication de priorité ou cité pour déterminer la date de publication d'une autre citation ou pour une raison spéciale (telle qu'indiquée) "O" document se référant à une divulgation orale, à un usage, à une exposition ou tous autres moyens "P" document publié avant la date de dépôt international, mais postérieurement à la date de priorité revendiquée "T" document ultérieur publié après la date de dépôt international ou la date de priorité et n'appartenant pas à l'état de la technique pertinent, mais cité pour comprendre le principe ou la théorie constituant la base de l'invention "X" document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme nouvelle ou comme impliquant une activité inventive par rapport au document considéré isolément "Y" document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme impliquant une activité inventive lorsque le document est associé à un ou plusieurs autres documents de même nature, cette combinaison étant évidente pour une personne du métier "&" document qui fait partie de la même famille de brevets		
Date à laquelle la recherche internationale a été effectivement achevée 28 mai 2021		Date d'expédition du présent rapport de recherche internationale 08/06/2021
Nom et adresse postale de l'administration chargée de la recherche internationale Office Européen des Brevets, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016		Fonctionnaire autorisé Kuntz, Jean-Marc

C(suite). DOCUMENTS CONSIDERES COMME PERTINENTS		
Catégorie*	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
A	EP 3 107 000 A2 (TATA CONSULTANCY SERVICES LTD [IN]) 21 décembre 2016 (2016-12-21) alinéa [0001] - alinéa [0009] alinéa [0018] - alinéa [0019] -----	4
A	XU ZEKUN ET AL: "Automatic Hyperparameter Tuning Method for Local Outlier Factor, with Applications to Anomaly Detection", 2019 IEEE INTERNATIONAL CONFERENCE ON BIG DATA (BIG DATA), IEEE, 9 décembre 2019 (2019-12-09), pages 4201-4207, XP033721319, DOI: 10.1109/BIGDATA47090.2019.9006151 le document en entier -----	1-3, 10-12
A	LIGHTFOOT LEONARD ET AL: "The effects of synthetically augmented training data on parameter tuning for anomaly detection algorithms", DEFENSE TRANSFORMATION AND NET-CENTRIC SYSTEMS 2011, SPIE, 1000 20TH ST. BELLINGHAM WA 98225-6705 USA, vol. 8062, no. 1, 13 mai 2011 (2011-05-13), pages 1-11, XP060014074, DOI: 10.1117/12.883526 le document en entier -----	1,10-12
A	ANTUNES ANTONIO LORVAO ET AL: "Adding Value to Sensor Data of Civil Engineering Structures: Automatic Outlier Detection", 2018 9TH INTERNATIONAL CONFERENCE ON INFORMATION, INTELLIGENCE, SYSTEMS AND APPLICATIONS (IISA), IEEE, 23 juillet 2018 (2018-07-23), pages 1-6, XP033511174, DOI: 10.1109/IISA.2018.8633586 le document en entier -----	1,10-12

RAPPORT DE RECHERCHE INTERNATIONALE

Renseignements relatifs aux membres de familles de brevets

Demande internationale n°

PCT/EP2021/056668

Document brevet cité au rapport de recherche		Date de publication	Membre(s) de la famille de brevet(s)	Date de publication
CN 110471946	A	19-11-2019	AUCUN	

EP 3107000	A2	21-12-2016	AU 2016204093 A1	12-01-2017
			CN 106257438 A	28-12-2016
			EP 3107000 A2	21-12-2016
			JP 6611677 B2	27-11-2019
			JP 2017010552 A	12-01-2017
			US 2016371228 A1	22-12-2016
