uTakeCare : la décentralisation des données personnelles pour un déconfinement respectueux dans le cadre du COVID-19 : vers une citoyenneté anonyme et numérique

Lamine Amour[1], Matthieu Quiniou[2], Sara Tucci[3], Hichem Bourak[4], Sami Souihi[1]

[1]) Laboratoire LISSI, Université de Paris-Est Créteil (UPEC), France & Edupi, France

[2]) UNESCO Chair "ITEN" (Fondation Maison des Sciences de l'Homme, University de Paris 8 - Vincennes – Saint-Denis, France) France

[3]) CEA-LIST, Gif-sur-Yvette, France

[4]) XHUMANISA & Banlieues Santé, ONG , France

lamine.amour@u-pec.fr, mquiniou@msh-paris.fr , sara.tucci@cea.fr, hichem.bourak@xhumanisa.org, sami.souihi@u-pec.fr

Résumé

Pour lutter contre le coronavirus (COVID-19), la plupart des pays ont opté pour une politique de confinement. Lorsqu'une décision de déconfinement doit être prise, une question se pose quant à la stratégie numérique à adopter : Faut-il opérer un suivi des citoyens ? Tous ou seulement les personnes ayant contracté le COVID-19 ? Faut-il prendre des mesures spécifiques pour protéger les personnes âgées ou les personnes souffrant de comorbidités ? De nombreuses applications et approches ont été proposées pour assurer la sécurité publique dans le cadre du COVID-19. Dans ce papier , nous commencerons par faire l'inventaire de ces applications, nous discuterons des stratégies et des technologies adoptées et nous les classerons par catégories. Par la suite, nous présenterons une approche consistant à calculer un score de vulnérabilité afin de proposer une solution pour protéger les personnes à risque. Ensuite, nous détaillerons l'architecture de "uTakeCare", une application open-source que nous avons mise en place, ainsi que la méthode utilisée pour calculer le score de vulnérabilité. Cette méthode est basée sur une théorie de la fonction de croyance et des techniques d'apprentissage automatique (ML, Machine Learning). Enfin, nous aborderons les questions éthiques et juridiques de cette application et les méthodes à utiliser pour en garantir le plein respect (par exemple, la preuve à divulgation nulle de connaissances, les contrats intelligents, etc.) afin de compléter les exigences du RGPD (Règlement général sur la protection des

données personnelles) par des solutions numériques et éthiques dès la conception (*ethic-by-design*).

mots-clés : COVID-19, Apprentissage automatique (ML), Preuve de connaissance zéro (ZKP), Blockchain, Éthique dès la conception, Sécurité.

2. INTRODUCTION

Selon l'Organisation mondiale de la santé (OMS), plus de 3 millions de cas de COVID-19 et plus de 200 000 décès ont été signalés à la fin du mois d'avril 2020[1]. Outre les cas confirmés, il existe également des cas suspects de COVID-19, dont la définition évolue au fil du temps et à mesure que l'épidémie se propage. En outre, les critères varient selon les pays.

Dans ce contexte, de nombreux pays ont opté pour une politique de confinement. Lorsqu'une décision de déconfinement doit être prise, une question se pose quant à la stratégie à adopter : Faut-il opérer un suivi de chaque citoyen ? Tous ou seulement les personnes qui ont contracté le COVID-19 ? Faut-il prendre des mesures pour protéger les personnes âgées ou les personnes souffrant de comorbidités ? De nombreuses applications et approches ont été proposées pour assurer la sécurité publique dans le cadre du COVID-19 [1, 2, 3, 4, 5, 6] mais plusieurs de ces applications présentent des risques élevés pour les droits et libertés fondamentaux.

L'objectif principal que nous présenterons dans ce papier est de proposer un dispositif numérique de santé publique suivant une approche ethic-by-design selon une approche résolument différente des projets existants généralement axés sur la recherche des contacts des personnes infectées. Notre proposition d'application numérique s'appelle "uTakeCare". Cette application n'est pas une application de suivi des personnes ayant contracté le COVID-19, mais une application qui contribue à assurer la distanciation sociale (notamment pour protéger les personnes les plus vulnérables). La philosophie de cette application repose sur la protection de la vie privée et l'identité auto-souveraine. L'approche n'est pas pseudonyme mais vise un anonymat complet bien au-delà des exigences du RGPD. Le score de vulnérabilité au COVID-19 est calculé à l'aide de techniques d'*apprentissage automatique(ML)* et de la *théorie des fonctions de croyance.* Pour l'utilisateur identifié comme

[1] https://covid19.who.int/

vulnérable, l'application demande à l'utilisateur de partager sa localisation (dans un processus totalement volontaire et anonyme) pour indiquer les endroits vulnérables (une carte "thermique" des zones vulnérables). La mise à jour de la localisation ne peut se faire qu'à la suite d'une action explicite de l'utilisateur. L'idée est d'alerter les utilisateurs qui passent par ces zones afin de s'éviter et de prendre conscience de la distanciation sociale. En outre, un utilisateur vulnérable peut activer le mode balise pour diffuser via Bluetooth une preuve anonyme permettant la mise en place de zones d'exclusion volontaire et citoyenne.

Le papier est structuré comme suit. La section 3 traite des principales applications en matière de sécurité publique. La section 4 présente une discussion éthique et juridique sur les applications numériques COVID-19. La section 5 décrit notre principale contribution intitulée: "uTakeCare". La section 6 montre les limites, les perspectives et les travaux futurs envisagés pour améliorer le dispositif. Enfin, la section 7 tire des conclusions.

3. APPLICATIONS DE SANTÉ PUBLIQUE COVID-19

Plusieurs applications ont été développées pour limiter l'impact du Covid-19 (tableau 1). Selon les législations et les pays, ces applications respectent plus ou moins la vie privée. Ces applications sont plus ou moins complexes. Certaines se limitent à fournir des informations, tandis que d'autres permettent de les personnes. Même ces dernières peuvent être classées en 2 catégories, celles basées sur l'utilisation volontaire et celles rendues obligatoires.

L'application la plus populaire est probablement "Alipay Health Code" [1]. Développée par le géant Alibaba, elle a été déployée en Chine. Elle permet à l'utilisateur d'obtenir un QR-code de couleur en fonction de ses déplacements (avec le GPS de son téléphone). Le QR code est en vert pour les personnes considérées sans risque, et aucune restriction de voyage ne leur est imposée. Lorsque le QR code est en jaune, la personne se voit imposer un confinement de 7 jours. Lorsque le QR code est en rouge, la personne se voit imposer un confinement de 14 jours.

En Corée du Sud, plusieurs applications ont été développées pour localiser les endroits les plus infectés. L'une des applications coréennes les plus populaires est l'application "Corona

100m" [2]. Elle est conçue pour aider les gens à reprendre une vie sociale prudente et à minimiser la propagation de la maladie en avertissant les utilisateurs de la présence possible de porteurs de la maladie dans une zone de 100 m. Cette application est également basée sur le GPS intégré dans le smartphone et sur une base de données centralisée.

En France, le gouvernement envisage de déployer une application appelée "Stop Covid" [3]. Cette application ne devrait pas permettre de suivre les personnes exposées au virus ou les diagnostics positifs. L'idée est de mettre en place une solution de « traçage des contacts » qui serait basée sur l'utilisation de la technologie « Bluetooth » pour construire une liste anonyme de contacts entre les smartphones. A terme, si l'un des contacts est positif, l'utilisateur en sera informé. Elle ne serait utilisée que sur une base volontaire et serait basée sur le protocole ROBERT, basé lui-même sur des données centralisées [4].

Récemment, Google et Apple ont également uni leurs forces pour offrir une solution de recherche de contacts, appelée «notification d'exposition», basée sur la technologie Bluetooth [5]. Cette application est également basée sur la participation volontaire. Si une personne est diagnostiquée positive, elle pourra alors entrer les résultats de ses tests dans l'application. Avec son consentement, ses clés d'identification anonymes seront ensuite envoyées à un serveur et stockées pendant 14 jours. L'idée est d'avertir l'utilisateur en cas de présence de personne dans son périmètre Bluetooth positive au COVID-19, sans préciser qui afin de préserver son anonymat. Outre l'approche douteuse de la centralisation des données (même pour 14 jours), cette application peut poser des problèmes dans une zone peu peuplée. Néanmoins, et malgré les inconvénients, plusieurs pays (par exemple, l'Allemagne) ont adopté la solution.

Le protocole DP3T (*Decentralized Privacy-Preserving Proximity Tracing*) pourrait être la meilleure solution disponible jusqu'à présent en terme de protection de la vie privée [6]. Il s'agit de la combinaison d'une approche centralisée et décentralisée. Tout d'abord, chaque smartphone conserve une trace des contacts avec les autres smartphones. L'anonymat est assuré par des fonctions de hachage et des fonctions de touches temporaires. Ensuite, dès qu'une personne est détectée comme positive, les clés générées sont entrées sur un serveur central. Chaque smartphone utilise ces informations pour estimer l'exposition au virus sur la

base des interactions passées. Toutefois, cette application n'est utile que si les cas positifs sont effectivement signalés.

Avec son application "Trace Together" [23], Singapour a inspiré de nombreux États à travers le monde tels que l'Australie (CovidSafe) [32], la Colombie (CoronApp) [33] ou le Sénégal (Covid-Trace) [25]. Utilisant le Bluetooth, lorsqu'il est activé sur les téléphones, elle permet de détecter autour de l'utilisateur d'autres téléphones ou d'autres appareils connectés. L'objectif est d'identifier les patients et les contaminations potentielles. À Singapour, cette application permet donc à une personne de savoir si elle a rencontré une personne infectée. Elle est même rétroactive : si une personne est testée positive a posteriori, les personnes qui l'avaient croisée quelques jours auparavant sont averties, invitées à rester chez elles, à se déclarer par téléphone si elles ressentent des symptômes et/ou à effectuer un test. À Singapour, les données doivent être effacées dans un délai de 21 jours, conformément à l'objectif principal de l'application qui est d'identifier et de circonscrire rapidement les foyers de contagion.

Touchée de manière différée par la propagation de la pandémie, de nombreux pays du continent africain optent également pour des applications de recherche et suivi des contacts.

L'application de traçage GH-Covid 19 du Ghana repose, par exemple, sur le principe du volontariat tout en demandant à l'utilisateur d'activer la localisation GPS [34].

Le Sénégal (Covid-Trace) a été inspiré par Singapour et son application de recherche des contacts, "Trace Together". Toutefois, il est trop tôt pour évaluer le déploiement de l'application sénégalaise. En effet, il est communément admis par les experts pour que ce type d'approche soit pleinement efficace, qu'un soutien massif de la population est nécessaire.

Au Maroc [26], en attendant le déploiement de la solution de recherche des contacts, qui serait sélectionnée fin avril par les ministres de l'Intérieur et de la Santé en coordination avec l'Agence de développement numérique, la Direction générale de la sécurité nationale utilise une solution plus coercitive.

En Côte d'Ivoire[24], le collectif Anticoro avec le dispositif "Pass Santé Mousso", dispose notamment d'une fonctionnalité de géolocalisation des utilisateurs qui émettent des demandes et qui sont estimés à risque. A noter l'approche inclusive louable et complémentaire illustrée par l'accès de l'utilisateur à l'ensemble des recommandations et mesures de prévention en 11 langues locales ainsi qu'aux numéros d'urgence. L'utilisateur peut bénéficier d'une assistance psychologique ou vidéo. Les programmes d'apprentissage en ligne et le télétravail confirment cette approche, qui est soutenue par la société civile en bonne intelligence.

Cependant, plusieurs ministères de la santé en Afrique préfèrent le DHIS2 (*District Health Information Software*), qui a publié une boîte à outils en réponse au Covid-19 [35]. Le DHIS2 est une plateforme HMIS (*Health Information Management System*) utilisée par les ministères de la santé de 72 pays à faible et moyen revenu.

L'approche de Taïwan en matière de sécurité sanitaire reste néanmoins collaborative et civique [17]. Bien que le traçage ne se fasse pas sur la base du volontariat, la majorité de la population l'a néanmoins accepté grâce à la transparence et à un appel à contribution des autorités auprès des citoyens. La société civile a proposé un grand nombre d'applications et de solutions, dont certaines ont littéralement remplacé les solutions développées par les autorités. À Taïwan, les « Hackers citoyens» luttent contre le Covid 19.Taïwan a fait en sorte que sa population accepte sa stratégie. La dimension culturelle est importante. Cependant, l'inclusion des citoyens semble se positionner, aux côtés de la sécurité et du respect de la protection des données personnelles, comme l'un des facteurs clés d'un système efficace de recherche des contacts entre citoyens.

Une approche inclusive pour l'ensemble de la population semble essentielle pour assurer la durabilité de l'efficacité de la stratégie adoptée et ainsi pour que la recherche des contacts soit pleinement efficace. La recrudescence du nombre de personnes infectées à Singapour est symptomatique de ce besoin de durabilité et d'acceptabilité du dispositif numérique **[36]**. Passant de 200 cas le 15 mars 2020 à 14 423 cas de personnes infectées le 29 avril 2020, les autorités sanitaires ont remis en cause leur méthode en renonçant à une forme de solutionnisme technologique **[37]**. Elles ont ainsi réalisé qu'elles avaient exclu de leur « gestion 4P » de la crise les populations immigrées, dont beaucoup vivent entassées dans des hangars et autres dortoirs dans des conditions précaires. Ses populations dites « invisibles »

concentrent la grande majorité (95%) des nouveaux cas. L'inclusion citoyenne doit prendre en compte les populations étrangères et les personnes en situation irrégulière comme les réfugiés et autres immigrants illégaux. Ils n'ont pas de papiers d'identité en règle mais sont actifs sur le territoire et, malheureusement, pour eux et pour toute la population, représentent un vecteur potentiel du virus. Singapour a dû réorienter sa stratégie sur une ligne plus dure et plus directive en imposant l'utilisation de l'application du QR Code (*SafeEntry)* **nécessaire** pouvoir entrer dans les lieux publics**[38]**.

Comme l'Inde et son application vedette Aarogya Setu qui a battu le record mondial de téléchargements en 13 jours avec plus de 50 millions **[39]**. Ce chiffre est impressionnant mais doit être relativisé par rapport à la population indienne qui compte 1,353 milliard d'habitants. Initialement téléchargeable sur une base volontaire, son téléchargement est maintenant obligatoire **[40]**.

Enfin, ces solutions de recherche des contacts ont été développées dans un contexte d'urgence, et la question de leur sûreté et de leur sécurité doit être appréciée avec précision et de manière concrète [41, 42, 43].

Pays	Nom de l'application	Suivi GPS	Recherche des contacts	Participation volontaire	Collecte centralisée des données
Chine	Code de santé d'Alipay	Oui	Oui	Non	Oui
Corée du Sud	Corona 100m	Oui	Oui	Oui	Oui
France	Stop Covid (ROBERT)	Non	Oui	Oui	Oui
États-Unis / Allemagne /...	Notification d'exposition	Non	Oui	Oui	Oui
Suisse	DP3T	Non	Oui	Oui	Les deux
Singapour	Trace Together SafeEntry	Non Non	Oui Oui	Oui Non	Oui Oui
Taïwan	Notification d'exposition	Oui	Oui	Non	Oui

Sénégal	Covid-Trace	Oui	Oui	Oui	Oui
Côte d'Ivoire[2]	Le collectif Anticoro avec le Pass Santé Mousso	Oui	Oui	Oui	Oui
Maroc	Notification d'exposition	NC[3]	Oui	Oui	Oui
Australie	CovidSafe	Non	Oui	Oui	Oui
Ghana	Application de suivi GH-Covid19	Oui	Oui	Oui	Oui
Inde	Aarogya Setu	Oui	Oui	Oui/Non. Récemment, elle est devenue obligatoire.	Oui
Colombie	CoronApp	Non	Oui	Oui	Oui
Plusieurs pays	DHIS2	Oui	Oui	Oui	Oui

Tableau 1. Applications développées pour le Covid-19 Sécurité publique

4- DISCUSSION ÉTHIQUE ET JURIDIQUE SUR LES APPLICATIONS NUMÉRIQUES DU COVID-19

La lutte contre une pandémie par les outils numériques est une situation sans précédent qui nécessite de trouver un équilibre entre la gestion optimale d'une crise sanitaire mondiale, ayant d'ailleurs conduit certains pays à confiner leur population, et la protection des données personnelles les plus sensibles, les données de santé.

Le RGPD *(Règlement général sur la protection des données)*, qui figure au niveau international comme un standard élevé pour la protection des données personnelles, n'est pas un outil conçu pour freiner une politique d'État intrusive sur la vie privée et le traitement des données personnelles.

[2] En Côte d'Ivoire, la solution Mediclick, qui est basée sur le principe du volontariat, n'a pas été retenue.

[3] Le choix de la solution à la fin du mois d'avril

Les bases légales du traitement sont énoncées à l'article 6 du RGPD. Elles sont plus ou moins rigoureuses pour un État, un organisme de santé ou une entreprise :

> *« 1. e traitement n'est licite que si, et dans la mesure où, au moins une des conditions suivantes est remplie : :*
> *a) la personne concernée a consenti au traitement de ses données à caractère personnel pour une ou plusieurs finalités spécifiques ;*
> *(...)*
> *d) le traitement est nécessaire à la sauvegarde des intérêts vitaux de la personne concernée ou d'une autre personne physique ;*
> *e) le traitement est nécessaire à l'exécution d'une mission d'intérêt public ou relevant de l'exercice de l'autorité publique dont est investi le responsable du traitement ; (...) ».*

Dans le cas d'un soutien public, il sera possible de se référer à une base légale (e) pour un organisme de santé (d) et pour une entreprise (a). Pour une entreprise souhaitant procéder à un traitement des données personnelles le consentement est nécessaire, il est donc préférable de bénéficier du soutien d'une autorité publique ou que celle-ci soit intégrée dans un processus mené par un organisme de santé.

En outre, comme indiqué ci-dessus, les données relatives à la santé sont des données particulièrement sensibles dont le traitement est en principe interdit en vertu de l'article 9 du RGPD :

> *« 1. Le traitement des données à caractère personnel qui révèle l'origine raciale ou ethnique, les opinions politiques, les convictions religieuses ou philosophiques ou l'appartenance syndicale, ainsi que le traitement des données génétiques, des données biométriques aux fins d'identifier une personne physique de manière unique, des données concernant la santé ou des données concernant la vie sexuelle ou l'orientation sexuelle d'une personne physique sont interdits.*
>
> *2. Le paragraphe 1 ne s'applique pas si l'une des conditions suivantes est remplie :*
>
> *(...)*
>
> *i) le traitement est nécessaire pour des motifs d'intérêt public dans le domaine de la santé publique, tels que la protection contre les menaces transfrontalières graves*

pesant sur la santé, ou aux fins de garantir des normes élevées de qualité et de sécurité des soins de santé et des médicaments ou des dispositifs médicaux, sur la base du droit de l'Union ou du droit de l'État membre qui prévoit des mesures appropriées et spécifiques pour la sauvegarde des droits et libertés de la personne concernée, notamment le secret professionnel ;.

En ce qui concerne les données de santé, génétiques ou biométriques, il convient également de noter que ce même article prévoit au point 4 une marge supplémentaire pour le pouvoir législatif national : « Les États membres peuvent maintenir ou introduire des conditions supplémentaires, y compris des limitations, en ce qui concerne le traitement des données génétiques, des données biométriques ou des données concernant la santé.

La seule voie ouverte par le RGPD pour une entreprise ne bénéficiant pas du soutien d'un Etat membre et souhaitant déployer une application avec des données très sensibles comme les données de santé est de s'appuyer sur l'article 11, c'est-à-dire l'anonymisation. La cryptographie, les blockchains publiques et la preuve à divulgation nulle de connaissance (*Zero Knowledge Proof*) peuvent être utiles et apporter à l'utilisateur plus de confidentialité que toute application numérique centralisée reposant sur des bases légales ouvertes aux autorités publiques pour traiter des données sensibles. La preuve à divulgation nulle de connaissance est le concept structurant de l'identité auto-souveraine.

Si les auteurs du RGPD sont souvent critiqués pour ne pas avoir pris en compte la technologie des chaînes de blocs (*blockchain*), notamment en ce qui concerne le droit à l'effacement (« droit à l'oubli), cette considération n'a pas d'importance tant que cette technologie est basée sur de solides systèmes de preuve à divulgation nulle de connaissances permettant une réelle anonymisation. Avec de telles garanties éthiques et techniques, le RGPD n'a pas besoin de s'appliquer et, selon son considérant 26, ne s'applique pas : « Il n'y a dès lors pas lieu d'appliquer les principes relatifs à la protection des données aux informations anonymes, à savoir les informations ne concernant pas une personne physique identifiée ou identifiable, ni aux données à caractère personnel rendues anonymes de telle manière que la personne concernée ne soit pas ou plus identifiable »**"**.

5. UTAKECARE : UN NOUVEAU CONCEPT D'APPLICATION DE DÉCONFINEMENT

5.1. IDÉE GÉNÉRALE

L'étude comparative des applications que nous avons menée dans la section 3 confirme que les applications d'intérêt public développées dans la lutte contre le coronavirus *(COVID-19)* souffrent d'inconvénients et ne sont pas orientées vers l'identité auto-souveraine. Cette analyse a permi de constater que chaque application est dédiée soit à la prédiction du risque associé au virus COVID-19, soit à l'enregistrement des personnes qui sont en contact avec ce virus, soit encore au suivi des déplacements des personnes sans respecter la vie privée. Dans ce contexte, l'approche proposée dans cet article se distingue nettement de l'existant et des travaux en cours dont nous avons connaissance. En effet, l'application « *uTakeCare* » n'est pas une application de traçage des personnes affectées par le COVID-19, mais une application qui permet d'assurer une distanciation sociale afin de protéger les personnes vulnérables. Plus précisément, « *uTakeCare* » a deux objectifs : l'évaluation automatisée de la vulnérabilité au COVID-19 et l'organisation d'un mécanisme de distanciation sociale orienté vers la protection des personnes à risque. La philosophie est résolument orientée vers la protection de la vie privée et l'identité auto-souveraine. L'approche n'est pas pseudonyme mais vise une anonymisation complète au-delà même des exigences du RGPD *(Règlement général sur la protection des données)*.

L'objectif de l'application « *uTakeCare* » est d'atteindre des résultats proches de ceux du confinement, à savoir limiter la propagation du virus, mais en regagnant la liberté de mouvement sans pour autant renoncer à la protection de la vie privée, c'est-à-dire sans la révélation de données sensibles de santé. Un déconfinement sous veille ou surveillance numérique répond également à un objectif de santé publique dans la mesure où le confinement rend plus difficile l'accès à certains soins, même essentiels, et crée ou accentue pour certaines personnes des troubles psychologiques. Par ailleurs, dans certains pays ou territoires (notamment l'Inde et plusieurs pays d'Afrique subsaharienne, etc.), la situation de confinement a des conséquences sur l'accès aux biens de première nécessité. L'extension du confinement doit donc également être discutée sous ce prisme sanitaire dans l'hypothèse d'une solution numérique efficace permettant d'obtenir les mêmes effets avec une plus grande

liberté et un plus grand accès aux produits essentiels. Il faut également noter que ce type de systèmes limiterait l'impact négatif de la gestion de cette crise sur l'économie en général.

Cette application est conçue pour être totalement anonyme au niveau du logiciel et du smartphone. Le résultat de l'évaluation de la vulnérabilité est chiffré avec un système de hachage cryptographique. Les données personnelles sensibles qui ont permis ce traitement ne sont effacées du cache de l'appareil automatiquement (smartphone). Avec cette approche, il est possible d'utiliser la preuve à divulgation nulle de connaissance *(Zero Knowledge Proof*[4]*)* pour permettre aux utilisateurs de prouver la vulnérabilité sans révéler ou même stocker des données personnelles. Il sera également proposé dans une version ultérieure de *«uTakeCare* » de renforcer la sécurité en ce qui concerne les appareils (smartphone, tablettes, etc.), avec un système de hachage du résultat de la vulnérabilité en mode déconnecté et un balayage de sécurité contre d'éventuels logiciels espions préinstallés (voir section (6)). La Figure (1) présente la description générale de « *uTakeCare* ».

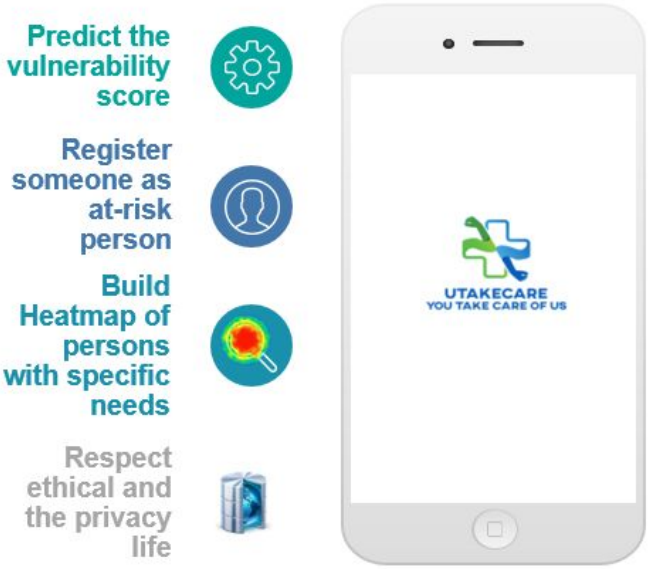

Figure 1. Description générale de l'application "uTakeCare".

Pour évaluer la vulnérabilité au COVID-19, nous mettons en œuvre un nouveau modèle d'estimation basé sur les techniques d'*apprentissage automatique*(*machine learning (ML))* et

[4] Matthieu Quiniou & Christophe Debonneuil, Blockchain Glossaire de l'UNESCO, 2019 "Zero Knowledge Proof (ZKP) consiste à prouver la détention d'une information sans en divulguer le contenu. Par exemple, prouver qu'une personne est majeure sans donner son âge est une Preuve de Connaissance Zéro. Le ZKP est largement utilisé car il suffit souvent de savoir que l'information est valable, sans qu'il soit nécessaire de la divulguer".

la *théorie des fonctions de croyance* (*théorie de Dempster-Shafer)*. Cette estimation est basée sur les informations fournies par un utilisateur, telles que les données de morbidité (par exemple, *hypertension, diabète,* etc.), l'âge, le poids et la taille. La Figure (2) montre les étapes par lesquelles l'utilisateur passe pour obtenir son score de vulnérabilité.

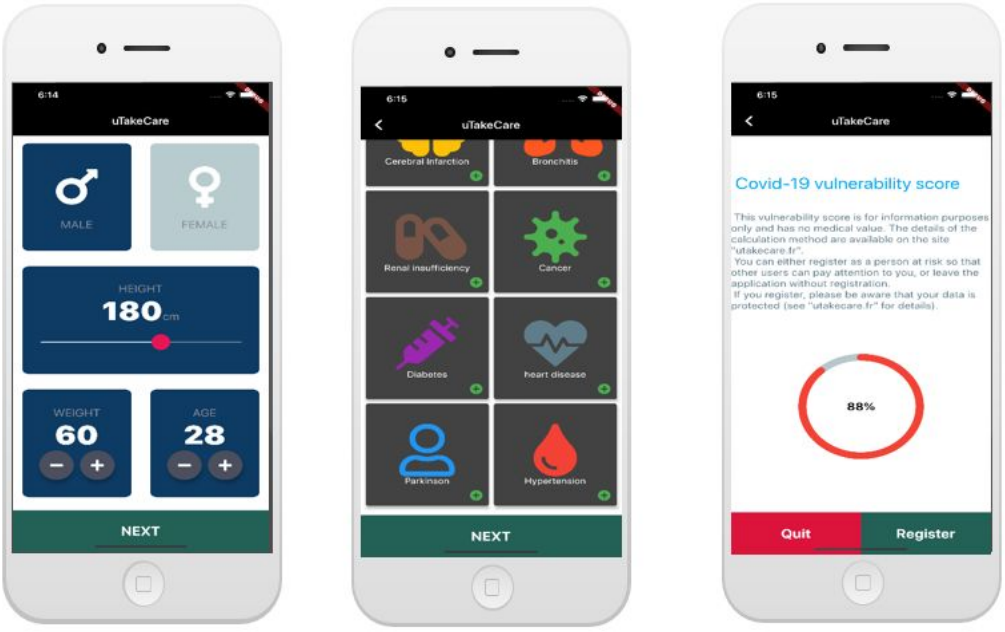

Figure 2. Présentation des interfaces "uTakeCare

En plus du calcul du score de vulnérabilité, l'application permet aux utilisateurs de déclarer leur localisation. L'objectif est de construire une carte de chaleur des zones vulnérables. La mise à jour de la localisation ne peut se faire qu'à la suite d'une action explicite de l'utilisateur. L'idée est d'alerter les utilisateurs qui passent par ces zones pour l'éviter ou pour prendre conscience de la distanciation sociale à respecter dans une situation particulière.

Enfin, un utilisateur vulnérable peut activer le mode balise pour diffuser via *Bluetooth* une preuve anonyme permettant de vérifier qu'il s'agit bien d'une personne vulnérable sans en connaître les détails. Par ce moyen, il est permis d'envisager de mettre en place des zones d'exclusion volontaire.

5.2. MÉTHODES EXISTANTES POUR L'ESTIMATION DE LA VULNÉRABILITÉ AU COVID-19

L'analyse des données collectées par les hôpitaux et distribuées sous forme de bases de données ouvertes ou semi-ouvertes offre un réel potentiel pour améliorer la qualité de vie et assurer la sécurité des populations dans la période post-confinement. Cependant, les outils actuellement à la disposition des citoyens ne sont pas suffisants ou sont mal adaptés à une utilisation et une analyse efficaces [13]. Cela est dû, d'une part, au fait que les informations proviennent de différentes sources d'information indépendantes utilisant chacune son propre modèle de données et, d'autre part, au fait qu'il n'existe aucun outil permettant de modéliser, de représenter ou d'utiliser ces données de manière simple et robuste pour estimer la vulnérabilité au COVID-19. Outre le manque de données disponibles, la performance des applications d'évaluation COVID-19 dépend également d'un certain nombre de paramètres, notamment l'âge, le contact avec les personnes infectées, les voyages dans les zones touchées ou même la morbidité d'un individu [9].

Deux types d'approches ont été développés pour évaluer le score de vulnérabilité COVID-19: l'approche classique de modélisation statistique et l'approche de modélisation par *apprentissage automatique* connu sous l'appellation anglo saxonne *machine learning (ML) (*Figure 3). Dans cette section, nous en présentons quelques-unes.

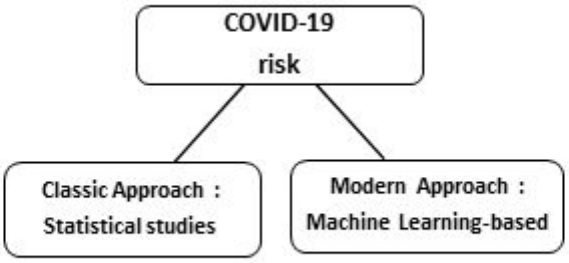

Figure 3 : évaluation de la vulnérabilité selon COVID-19.

La première étude que nous présentons est l'étude [8]. Chen et al. se sont basés dans celle-ci sur des informations recueillies auprès de 99 patients atteints du COVID-19 dans la *province de Hubei en Chine.* Dans ce travail, l'analyse statistique est utilisée pour décrire les données épidémiologiques, les signes et les symptômes des patients à l'admission, leur comorbidité, etc. Dans cette étude, les auteurs utilisent des mesures statistiques pour présenter la relation

entre les symptômes et les problèmes respiratoires graves. Ce travail fournit de nombreux résultats intéressants sur les symptômes. Un autre résultat intéressant est que les hommes âgés présentant des comorbidités sont plus fréquemment atteints de graves problèmes respiratoires, qui dans certains cas sont mortels.

La deuxième étude que nous présentons consiste dans l'article [10]. Dans cette étude, Wang et al. se sont basés sur les informations recueillies auprès de 165 patients atteints du COVID-19 dans la province de *Fujian en Chine*. Ce travail compare les résultats obtenus avec la conclusion tirée sur la base des informations recueillies à *Wuhan* (*province de Hubei*). Une analyse statistique est également utilisée pour déterminer les caractéristiques épidémiologiques et cliniques des patients atteints de la maladie COVID-19. Selon les auteurs, cette étude suggère que la plupart des patients atteints de COVID-19 de la province du Fujian ne sont pas des cas graves et que le taux de mortalité est extrêmement faible, ce qui est similaire à celui des autres provinces de Chine, à l'exception de la province du Hubei. Les symptômes au début de la maladie sont principalement la fièvre (76,4 %), la toux (60 %) et l'expectoration (38,2 %). Cette étude confirme que les patients âgés ont tendance à présenter davantage de comorbidités comme la lymphopénie, l'hypoxémie et une longue période d'excrétion virale. Cela pourrait expliquer pourquoi ils sont plus susceptibles d'évoluer vers une forme grave ou critique et ont un risque de décès plus élevé que les patients plus jeunes.

Dans une autre étude [9], Adlhoch et al. proposent un rapport européen utilisant l'analyse statistique pour prédire le risque de cas graves nécessitant une assistance respiratoire. Cette étude explore un large panel de données (informations recueillies sur 58 277 cas dans 11 pays *e*uropéens (UE) et de l'Espace économique européen (EEE)). Selon ce travail, 32 % des cas diagnostiqués ont nécessité une hospitalisation et 2,4 % ont une maladie grave nécessitant une assistance respiratoire. En outre, la probabilité d'hospitalisation, de maladie grave et de décès augmente chez les personnes de plus de 70 ans et celles qui souffrent de maladies sous-jacentes telles que l'hypertension, le diabète, les maladies cardiovasculaires, les maladies respiratoires chroniques, l'état immunodéprimé, le cancer et l'obésité (73,4 % des patients gravement malades avaient un indice de masse corporelle (IMC) de plus de [30-40]).

Dans une autre étude [16], Garg et al. utilisent l'analyse statistique pour produire des taux d'hospitalisation hebdomadaires associés au COVID-19, classés par âge. Les auteurs

suggèrent que les hospitalisations associées au COVID-19 aux États-Unis d'*Amérique (USA)* sont les plus élevées chez les adultes âgés, et que près de 90 % des personnes hospitalisées présentent une ou plusieurs pathologies sous-jacentes. En fait, ils confirment l'importance des mesures préventives (par exemple, l'éloignement social, l'hygiène respiratoire et le port des masques lorsque les mesures d'éloignement social sont difficiles à maintenir) pour protéger les personnes âgées et les personnes souffrant de conditions médicales sous-jacentes. L'étude a également noté que 54 % des patients hospitalisés en raison du COVID-19 sont des hommes et 46 % des femmes.

La première étude qui utilise les méthodes *ML* pour évaluer le COVID-19 est présentée dans l'article [12]. Dans ce travail, Jia et al. adoptent trois types de modèles *ML*, à savoir, le *modèle logistique, le modèle de Bertalanffy* et *le modèle de Gompertz* pour évaluer le risque COVID-19. Les tendances épidémiques de la maladie *SRAS (Syndrome Respiratoire Aigu Sévère)* ont d'abord été analysées afin de prouver la validité des modèles mathématiques existants. Ensuite, les résultats ont été utilisés pour les adapter à la situation de la maladie du COVID-19. Les résultats des prédictions des trois modèles testés sont différents selon les paramètres et les régions analysées. Selon cette étude, le *modèle logistique* semble être la meilleure approche.

Dans l'étude [11] qui utilise les techniques de *ML*, DeCaprio et al. proposent des modèles basés sur le *ML* pour identifier les individus vulnérables (risque de mortalité). Un large ensemble de données, avec plus de trois millions d'échantillons, est utilisé. Trois modèles de *ML* (*régression logistique, arbres à gradient* et un vaste ensemble de caractéristiques générées à partir des données des demandes de remboursement de Medicare) ont été formés pour calculer le score de vulnérabilité COVID-19 d'une personne. D'après les résultats, les performances des modèles d'arbres à gradient renforcé sont supérieures aux autres. La fonction d'efficacité du récepteur, plus fréquemment désignée sous le terme « courbe ROC » (receiver operating characteristic), montre que lorsque le seuil de décision augmente, le pourcentage de la population potentiellement affectée augmente à peu près au même rythme. Ce travail donne également accès gratuitement aux modèles ML utilisés. Cela peut encourager la collaboration de la communauté des logiciels libres pour améliorer les modèles proposés.

L'idée de base de l'étude proposée par Dandekar et al [13] est de concevoir une méthodologie complète qui tente d'interpréter et d'extrapoler les données disponibles publiquement en utilisant un modèle mixte d'équations épidémiologiques de premier ordre et de *réseaux neuronaux artificiels (ANN)* pilotés par les données. Quatre ensembles de données provenant du *Wuhan*, de *l'Italie*, de la *Corée du Sud* et des *États-Unis d'Amérique (USA)* sont utilisés pour tester le modèle proposé. L'objectif est de comparer le rôle joué par les mesures de quarantaine et de confinement. D'après les résultats, les pays dans lesquels des interventions gouvernementales rapides et des mesures de santé publique strictes de quarantaine ont été mises en œuvre ont réussi à stopper la propagation de la pandémie et à empêcher son explosion exponentielle.

Dans l'étude [14], Pourhomayoun et al. proposent une méthode appropriée pour déterminer le risque sanitaire et prédire le risque de mortalité des patients atteints de COVID-19. Elle est basée sur des algorithmes utilisant un certain nombre de méthodes de *ML* utilisées séparément. Les auteurs ont analysé et comparé les performances de ces méthodes afin d'identifier les caractéristiques importantes pour prédire le risque de mortalité. Les méthodes utilisées sont les suivantes : *les machines à vecteurs de support (SVM), les réseaux neuronaux artificiels (ANN), les forêts aléatoires (RF), arbre de décision (DT), régression logistique et la méthode des k plus proches voisins (KNN)*. Dans l'évaluation des résultats, la méthode des réseaux neuronaux artificiels a été le meilleur modèle avec une précision de plus de 93,75%. Les auteurs expliquent que le modèle proposé aide les hôpitaux et les établissements médicaux à décider qui doit être soigné en premier, qui a une priorité plus élevée pour être hospitalisé.

Dans l'étude [15], Batista et al. comparent plusieurs méthodes de *ML* pour prédire le risque de diagnostic COVID-19 positif et les résultats des examens d'admission aux soins d'urgence. Cinq algorithmes d'apprentissage automatique (les *réseaux de neurones (RNN), les forêts aléatoires (RF), les arbres basés sur la méthode de boosting gradient, la régression logistique et les machines à vecteurs de support (SVM)*) ont été formés sur un échantillon aléatoire de 70 % des patients brésiliens. Selon les données, la méthode *SVM* est la plus efficace, avec une fonction d'*efficacité du récepteur (ROC)*, une *sensibilité* et un *score de brier de* 0,85, 0,68 et 0,16, respectivement. Alors que la méthode *RF* est la deuxième meilleure méthode pour calculer le risque de COVID-19 positif. Selon cette étude, les trois variables les plus

importantes pour la performance prédictive étaient le nombre de lymphocytes, de leucocytes et d'éosinophiles.

5.3. UTAKECARE ESTIMATEUR DE VULNÉRABILITÉ AU COVID-19

Cette section présentera un nouvel outil d'estimation du score de vulnérabilité COVID-19. Il s'appuiera sur des bases de données d'informations d'utilisateurs obtenues de plusieurs sources [9,16,20,21,22]. Par conséquent, nous avons accordé une attention particulière à la mise en œuvre d'un modèle de données unifié simple basé sur la théorie de fusion d'informations. Ce modèle devrait être ouvert, gratuit, évolutif et réutilisable dans de futurs projets et par la communauté.

Dans la section (5.2), nous avons présenté différentes approches d'estimation. Cependant, l'intégration de différentes sources nécessite un modèle mathématique complexe capable d'intégrer toutes les variables qui ont un impact sur la mortalité due au COVID-19. Celles-ci peuvent être des informations sur l'utilisateur, notamment l'âge, le sexe, le poids et la taille d'une personne ou même ses états de comorbidité comme le diabète, l'hypertension, les maladies cardiaques, la bronchite chronique, etc. Compte tenu de la diversité des données et des modèles prédictifs disponibles, il semble nécessaire de développer de nouvelles approches pouvant être adaptées à différents contextes afin de prédire efficacement le risque de mortalité des patients atteints du COVID-19. Le développement de ce type de méthode est exploratoire, et il est difficile de prévoir ses performances.

Notre nouveau modèle appelé « *uTakeCareCBF* » vise à résoudre ce problème. Il est basé sur l'utilisation de la théorie des fonctions de croyance (théorie de Dempster-Shafer) et des méthodes *ML*. L'idée clé est d'utiliser plusieurs estimateurs en même temps, afin d'ajuster l'estimation en fonction du contexte. En pratique, l'objectif du modèle « *uTakeCareCBF* » proposé est de calculer, en amont, la vulnérabilité d'une personne au COVID-19. Il est composé de deux étapes (Figure 4). La première consiste à former un nouvel estimateur basé sur les méthodes *ML* entraînées avec les trois bases de données COVID-19 existantes, et la seconde consiste à calculer le score de vulnérabilité en fusionnant les résultats de ces méthodes en utilisant la théorie des fonctions de croyance.

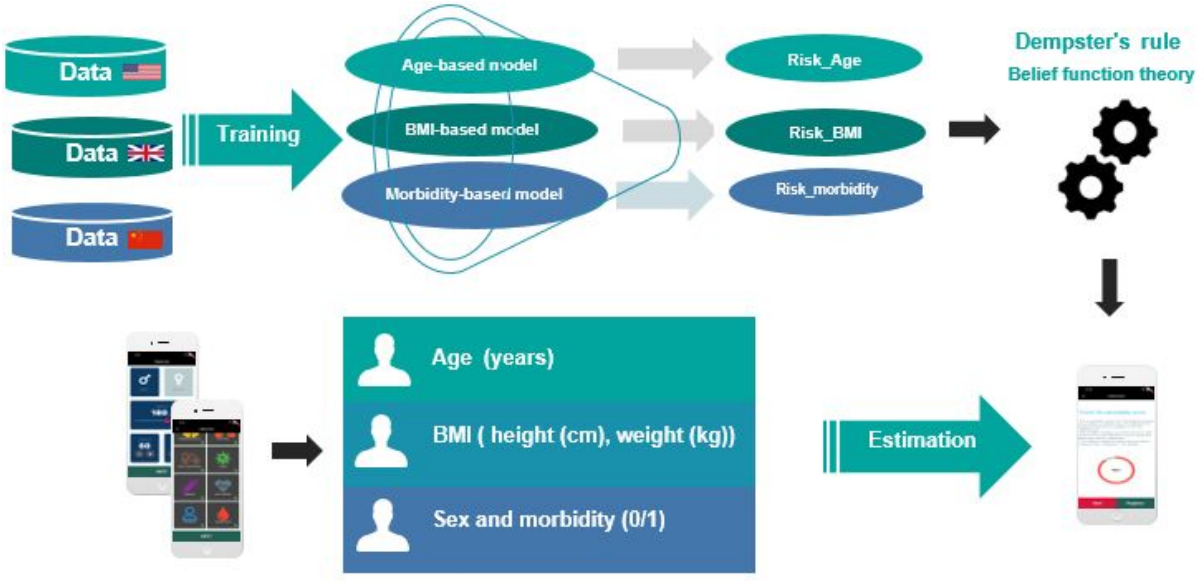

Figure 4. Présentation des étapes de l'*uTakeCareCBF*

5.3.a) Phase d'entrainement

La première étape du modèle « *uTakeCareCBF* » consiste en une phase de formation des modèles (entrainement). Dans cette étape, plusieurs ensembles de données [9,16,20,21,22] sont utilisés pour entraîner les méthodes choisies en fonction de l'âge, de l'indice de masse corporelle (IMC) et des informations sur la morbidité.

La première méthode consiste en un modèle linéaire basé sur l'âge et utilisant l'ensemble des données [16]. L'algorithme itératif *RANSAC* [28] est formé pour évaluer l'impact de l'âge sur la vulnérabilité au COVID-19. Les résultats de cette étude indiquent que la précision de l'algorithme RANSAC dans la détection du risque de mortalité dûe au COVID-19 en fonction de l'âge est d'environ 90%.

La deuxième méthode est le *modèle polynomial* basé sur l'IMC *(POLY)*. Celle-ci est basée sur [9] . Pour fixer le meilleur degré polynomial, un ensemble de valeurs est testé. Ce modèle atteint une précision de 87%.

La troisième méthode utilisée est la méthode de la *Random Forest (RF)* qui quantifie le risque de morbidité en utilisant des parties d'ensembles de données [20,21,22]. En effet, nous avons analysé et comparé les performances de plusieurs méthodes de *ML* qui ont contribué à prédire

le risque de mortalité comme cela a été fait dans les travaux [14, 15]. Les méthodes utilisées sont les suivantes : *Support Vector Machine (SVM), Artificial Neural Networks, Random Forest, Decision trees-based bagging et Gradient-based boosting*. La figure (5) donne un aperçu des facteurs considérés ainsi que leur corrélation avec la *valeur "r" de Pearson.* Une description détaillée de l'ensemble de données peut être trouvée dans le dépôt Gitlab [29].

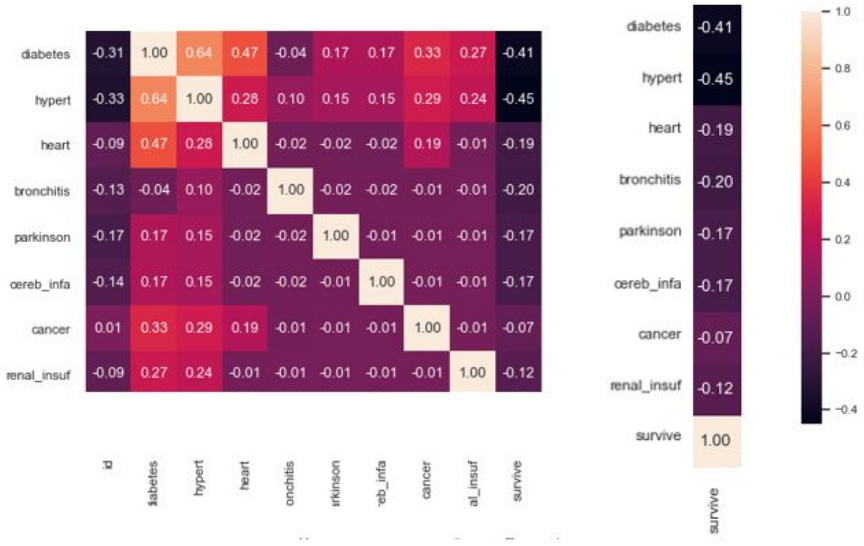

Figure 5. Matrice de corrélation ("r" Pearson).

Sur la base de cet ensemble de données, nous réalisons l'étape de validation croisée (10 fois) et l'étape d'ajustement des paramètres, au cours de laquelle diverses configurations sont ajustées pour sélectionner la combinaison de facteurs la plus optimale pour chaque méthode de *ML*. Pour ces deux étapes, nous avons suivi la même méthodologie que pour nos travaux précédents [31]. Pour plus de détails, veuillez consulter ce travail.

Les performances obtenues pour chaque méthode de ML sont présentées dans le tableau (2). Elles sont exprimées en termes d'erreur quadratique moyenne -*MSE*- (colonne 2), d'erreur absolue moyenne -*MAE*- *(*colonne 3) et de coefficient R au carré -r- (colonne 4).

Méthode	MSE	MAE	R au carré
Arbres aléatoires (RF)	**0.098**	**0.201**	**0.88**
Machines à vecteurs de support (SVM)	**0.110**	**0210**	**0.85**
Méthode bagging basé sur les arbres	**0.127**	**0.263**	**0.84**
Méthode boosting basée sur le gradient	**0.135**	**0.292**	**0.81**
Réseau de neurones (RNN)	**0.152**	**0.331**	**0.78**

Tableau 2. Performances obtenues

5.2.b) Étape de calcul

Sur la base des méthodes apprises (*RANSAC, POLY et RF*), les scores de vulnérabilité COVID-19 (risque lié à l'âge, risque lié à l'IMC et risque lié à la morbidité) sont calculés par le *modèle "uTakeCareCBF"*. Deux modèles basés sur *les forêts aléatoires (RF)* sont réalisés. Le premier est le modèle "*RF_female*". Il évalue le risque de morbidité au COVID-19 pour les personnes de sexe féminin. Le second ("*RF_male*") calcule le risque de morbidité au COVID-19 pour les personnes de sexe masculin, tel que présenté dans la figure (6).

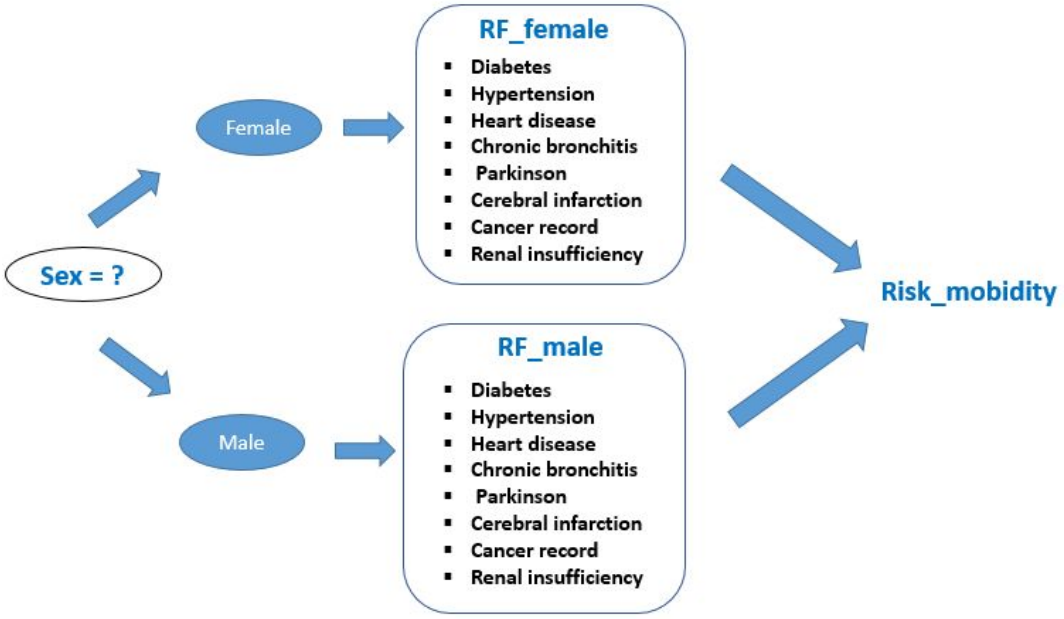

Figure 6. Présentation des *modèles RF réalisés*

Les scores de vulnérabilité calculés sont fusionnés pour déterminer la vulnérabilité globale au COVID-19 en utilisant la *théorie de la fonction de croyance*. En effet, le *«modèle de croyance transférable (TBM) » [30] est* mis en œuvre à cette fin. Les sources d'information, que nous appelons « oracles », sont les méthodes entraînées (RANSAC, POLY*, RF_female ou RF_male)*. Le modèle TBM est implémenté dans « uTakeCareCBF » *comme suit :*

- Le cadre de discernement du TBM est défini comme les scores de valeur de risque (*risk*$_i$) a*u niveau du crédal* du modèle TBM *[30]*. Ces scores sont la valeur du risque d'âge (*Risk_Age*), la valeur du risque d'IMC (*Risk_BMI)* et la valeur du risque de morbidité (*Risk_morbidity*).
- En utilisant les valeurs prédites par les oracles (*risk*$_i$), une approche heuristique est utilisée pour construire une fonction de masse (m) au niveau du crédal du modèle TBM. En résumé, pour chaque valeur de risque (*risk*$_i$), une valeur de masse initiale (m_i) est définie au départ. Ensuite, ces valeurs sont réparties sur les sous-ensembles du cadre de discernement [30].
- En utilisant les fonctions de masse calculées (m_i), la règle de Dempster's est mise en œuvre au *niveau pignistique du TBM*. L'étude [18] a montré comment ces fonctions sont fusionnées en calculant le coefficient de conflit entre différentes valeurs de risque (*Risk_Age*, *Risk_BMI* et *Risk_morbidity*). De nombreuses règles alternatives,

présentées dans l'étude [19], peuvent être utilisées pour redistribuer le coefficient de conflit. Enfin, la *crédibilité maximale (Bel)* est choisie au niveau pignistique du TBM pour quantifier le risque global au COVID-19.

Enfin, nous voulons souligner que le « *uTakeCareCBF* » est composé, pour l'instant, de trois parties comme présenté ci-dessus, mais il peut être facilement modulable et permettre l'interaction avec d'autres plateformes et de futures sources de données comme les médecins spécialisés, capables de reconnaître les risques potentiels. L'idée dans les travaux futurs serait d'intégrer les avis des médecins dans la phase de calcul comme source supplémentaire avec une plus grande confiance.

5.4. LA PREUVE À DIVULGATION NULLE DE CONNAISSANCE (ZKP): SOLUTION POUR OBTENIR UN ANONYMAT COMPLET ?

« uTakeCare » est conçu pour être totalement anonymisé au niveau du logiciel. Afin d'atteindre cet objectif, nous avons utilisé la technique preuve à divulgation nulle de connaissance ("*Zero-Knowledge Proof" (ZKP))*. L'idée principale est de permettre à l'utilisateur de prouver à une entité tierce qu'une proposition est vraie (dans notre cas, «qu'il s'agit bien d'une personne vulnérable ») sans révéler aucune autre information. Pour être plus précis, « uTakeCare » est basé sur une variante de ZKP qui ne nécessite pas d'interactions, appelée NiZKP (Non-Interactive *Zero-Knowledge Proof)* et utilise l'heuristique de Fiat-Shamir. Notre approche est très inspirée du *Zk-SNARK (Zero-knowledge succinct non interactive argument of knowledge)* [7]. Elle est composée de 3 éléments : Un générateur à base de graines, un générateur de preuves, et une fonction de vérification. Le circuit "C" est codé en dur dans l'application.

Le générateur à base de graines (figure 7) prend un paramètre secret, l'UUID (identificateur universel unique) du téléphone de l'utilisateur et un sel généré au hasard, et génère deux clés : *une clé de preuve (Key_P)* et une clé de *vérification (Key_V)*.

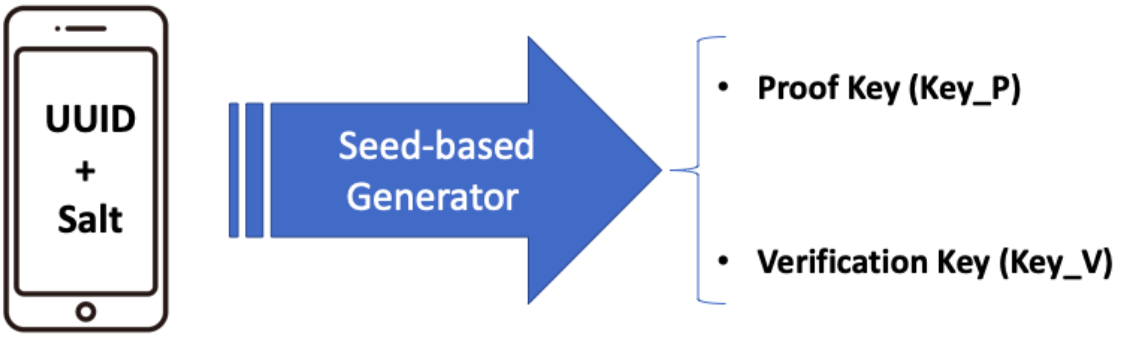

Figure 7. Description du générateur à base de semences

Le générateur de preuves (figure 8) prend la *clé de preuve (Key_P)* comme entrée ainsi que le score et le seuil de vulnérabilité calculés. Son travail consiste à générer la preuve. Comme le score de vulnérabilité ne change pas, il n'est autorisé à générer la preuve qu'une seule fois. Ainsi, *Salt* & Key_P, le score de vulnérabilité, doit être détruit après la génération de la preuve.

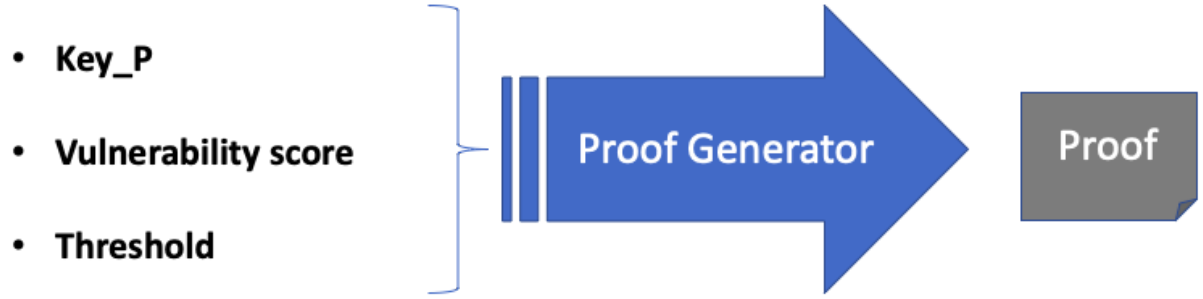

Figure 8. Générateur de preuves

L'étape de vérification (figure 9) prend la *clé de vérification (Key_V)* comme entrée ainsi que la preuve et le seuil. Son rôle est d'accepter ou de refuser la preuve.

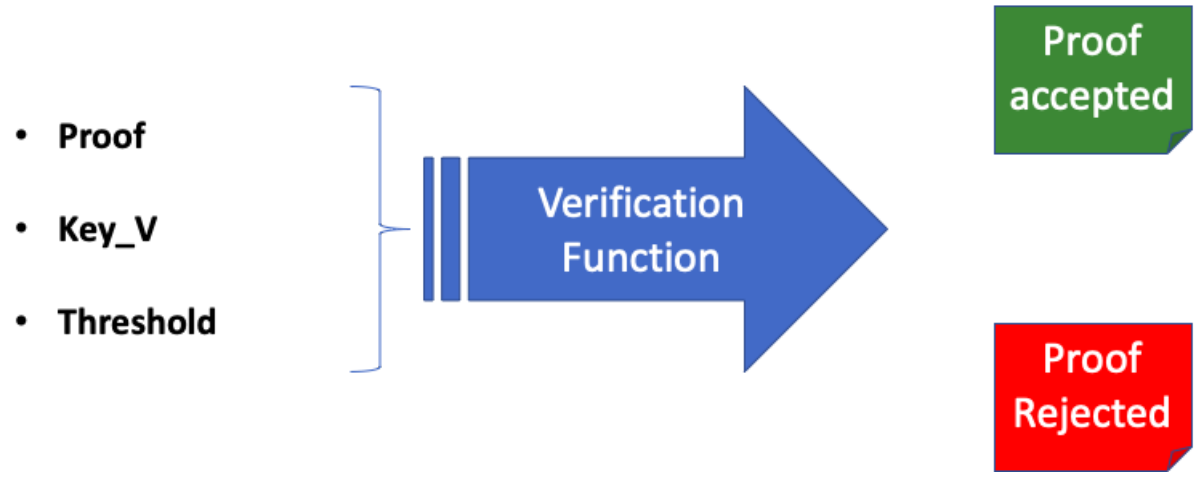

Figure 9. Fonction de vérification

Scénarios d'application

Pour l'utilisation de l'application, nous envisageons le protocole suivant, illustré à la figure 10. Dès que deux agents Alice et Bob sont connectés, Alice peut notifier à Bob son besoin de distance par le biais du message «la distance est justifiée ». Dès que Bob reçoit le message, il peut décider de vérifier la preuve (voir section 5.4) sur la base de son état de santé (connu de lui seul). Si Bob est vulnérable, il sera toujours d'accord sans même vérifier la preuve, car il a également intérêt à prendre un peu de distance avec Alice. Si Bob est en bonne santé, il peut éventuellement vérifier la preuve. La décision de vérifier ou non la preuve d'Alice est contextuelle : par exemple, s'il y a assez de place pour les deux, Bob n'a aucun intérêt à vérifier la preuve et il accordera immédiatement la distance à Alice par un message « ok ». Au contraire, s'il décide de vérifier la preuve, il n'accordera la distance que si la preuve est acceptée, sinon, aucun message n'est envoyé.

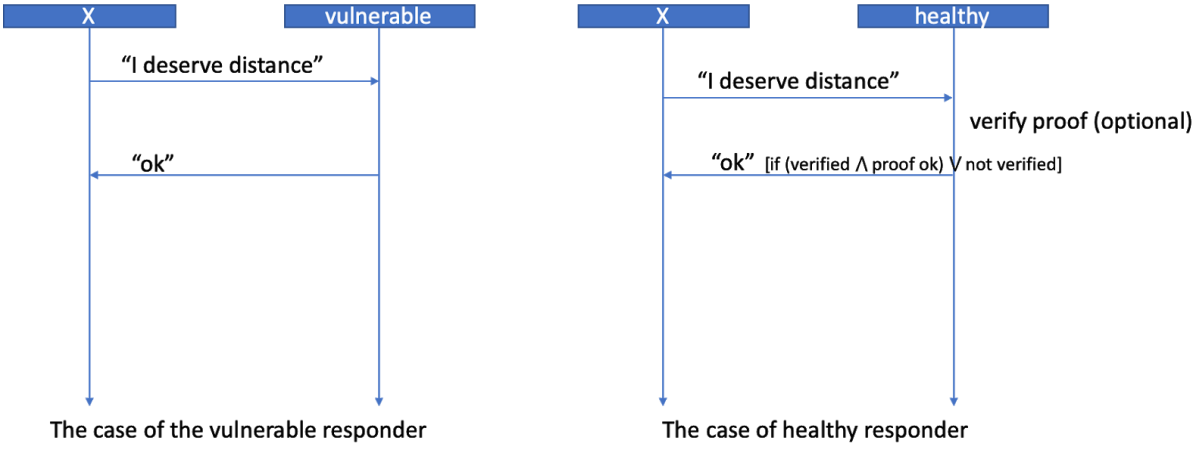

Figure 10. Scénarios d'application

Pour empêcher qu'un troisième agent ne devine l'état de santé d'Alice et de Bob en examinant leurs interactions, nous les rendons indiscernables en rendant la demande obligatoire pour tous les agents (tous les agents la demanderont et en cas de collision, le protocole redémarrera avec un mécanisme de retour en arrière) et en randomisant le comportement du répondant sain : il décidera de vérifier la preuve en tirant à pile ou face par exemple. Il est intéressant de noter que le même protocole pourrait être étendu pour inclure les personnes potentiellement positives au Covid-19, dans une future version de l'application, dans laquelle une preuve à divulgation nulle de connaissance pourra être produite pour cette catégorie de personnes. Dans ce cas, la personne positive se comportera comme le répondant vulnérable, et de plus, le vérificateur se contentera de tester les preuves sans savoir si la preuve concerne une personne vulnérable ou potentiellement positive au Covid-19. Le jeu des trois catégories apporte encore plus d'anonymat, puisque le répondant ne peut à aucun moment deviner l'état de santé de la personne qui demande.

6. LIMITES, PERSPECTIVES ET TRAVAUX FUTURS

Du point de vue de la sécurité, pour améliorer l'application «uTakeCare », plusieurs développements pourraient être envisagés dans le cadre d'une deuxième version de l'application, en particulier : un mode déconnecté guidé pour les phases de diagnostic et de

génération de ZKP ainsi qu'une chaîne de blocs (*blockchain*) publique qui stocke les clés de vérification.

La création d'un mode déconnecté pour les phases de diagnostic et de génération de ZKP permettrait de limiter les failles de sécurité liées à l'utilisation d'un smartphone connecté à Internet. Il serait ainsi envisagé de réaliser le diagnostic et le processus ZKP en mode déconnecté. Ce mécanisme permettrait de réaliser l'ensemble de l'opération impliquant des données personnelles sensibles en mode déconnecté.

Afin d'améliorer la robustesse de la solution, il serait préférable que le statut de «personne vulnérable » soit délivré par un médecin. On peut imaginer que dans le cas d'un score de vulnérabilité élevé, un médecin pourrait prendre le relais pour donner son avis. Pour éviter tout échange de données, le médecin peut fournir un code *QR (Quick Response) qui sera* scanné par l'utilisateur afin d'activer le générateur de preuves (section (5.4)). Le recours au médecin pour le diagnostic resterait volontaire et facultatif afin de permettre aux personnes ne souhaitant pas communiquer leurs données avec le médecin ou le système de santé de continuer à utiliser le dispositif.

En ce qui concerne le stockage des clés de vérification dans une chaîne de blocs, un tel mécanisme empêcherait toute altération unilatérale de la base de données, limitant ainsi le rôle des intermédiaires (serveur informatique de l'administration publique ou d'un hôpital, nuage souverain, etc.). Une architecture de chaîne de blocs légère serait préférable pour la latence, la sécurité et la mise à l'échelle. Une solution tolérante au problème des généraux byzantins, aux pannes, même sans commande totale des mises à jour, suffirait à garantir la cohérence des informations répliquées.

Une autre amélioration de la sécurité consisterait à utiliser un mécanisme de *bullet proof* au lieu de l'approche inspirée des Z-snarks décrite à la section 5.4 pour éviter la mise en place de Z-snarks de confiance. En effet, si la phase de génération de la clé est compromise, c'est-à-dire la valeur UUID connue par un logiciel malveillant, il serait possible de générer des preuves de tricherie acceptées par le vérificateur ; en outre, personne ne pourrait dire que la configuration de confiance a été subvertie. Les *bullet proof* conviennent toutefois aux preuves

de distance (comme dans notre cas), mais pour les vérifications plus complexes, les Z-Snarks offrent plus de souplesse.

Un autre point important pour apporter plus de sécurité, est de déclarer publiquement que toutes les données personnelles utilisées pour calculer le score de vulnérabilité, ainsi que le score de vulnérabilité et les preuves générées, ont été supprimées. Une approche intéressante pourrait consister à rendre l'utilisateur pleinement conscient de la suppression par le biais d'une interface appropriée demandant la suppression des données. Le résultat de l'opération (succès/échec) sera transcrit dans une blockchain publique. La notification de la transcription pourrait être visualisée sur l'application mobile une fois que la transaction de la blockchain associée soit validée.

L'utilisation d'une blockchain publique pour le stockage de la clé de vérification (ou d'un contrat intelligent de vérification si l'algorithme du vérificateur est rendu public) et l'enregistrement des opérations de suppression doivent être gratuits pour les utilisateurs, mais pour interagir avec les blockchains publiques, l'émetteur doit payer des frais. Une solution simple consisterait à faire appel à un intermédiaire, interposé entre l'utilisateur et la blockchain, pouvant émettre toutes les transactions au nom des utilisateurs et payer les frais associés. Cependant, cette solution rompt avec la décentralisation et la transparence. Par conséquent, pour éviter les intermédiaires, nous considérons l'existence d'un opérateur qui transférera par le biais des transactions de la blockchain la quantité nécessaire de crypto-monnaie aux portefeuilles des utilisateurs. L'application créera en effet automatiquement un portefeuille dans la blockchain cible associé à un pseudonyme dérivé de l'UUID du téléphone de l'utilisateur, par exemple. Les opérations de vérification et de suppression ayant lieu sur un téléphone donné seront signées par l'application de l'utilisateur et envoyées à la blockchin afin de garantir la traçabilité du journal des opérations.

Dans l'intention de généraliser l'adoption, deux autres approches méritent d'être étudiées. La possibilité pour certaines personnes, peu à l'aise avec la technologie, d'acquérir l'application déjà configurée par un professionnel médical certifié sur un smartphone conçu pour les seniors. On peut également imaginer un objet connecté (par exemple un bracelet non intrusif

prêt à être connecté par Bluetooth) avec des fonctions similaires, notamment un bouton à presser pour déclarer ponctuellement sa présence en tant que personne vulnérable.

- L'application uTakeCare pourrait également, à terme, être interopérable avec un chatbot garantissant l'inclusion des citoyens dans la solution. Le chatbot enverrait des alertes à différents moments de la journée pour demander à la personne vulnérable de prendre sa température ou de signaler des symptômes. La personne pourrait lui poser toutes sortes de questions afin de maintenir une certaine convivialité, même si elle est numérique artificielles.

Contrairement à l'approche actuellement proposée, qui repose sur des critères de risque déjà établis avec un haut niveau de fiabilité (âge, obésité, comorbidité, etc.), cette approche alternative repose sur des indicateurs qui ne sont pas encore fiables et qui pourraient être radicalement remis en question. Par exemple, en cas de mutation répétée du virus, la possibilité avérée de réinfection, la découverte de nouveaux facteurs de risque, etc. Cette approche mérite néanmoins d'être étudiée en raison de la pertinence de son utilisation en termes de conformité des entreprises (*compliance*), par exemple en termes de gestion des risques pour la santé et la sécurité des employés.

Parmi les autres considérations importantes à prendre en compte dans les versions futures figurent l'échelle et la limitation des effets indirects potentiellement discriminatoires.

Afin de limiter les effets potentiellement discriminatoires de ce type d'application numérique et d'ajouter une couche à la preuve à divulgation nulle de connaissance, une approche consisterait à passer de la preuve de vulnérabilité à la preuve d'incompatibilité ou à la preuve de dangerosité.

Dans un tel modèle, il y aurait trois catégories d'utilisateurs :

1) les personnes se déclarant peu vulnérables et a priori non contagieuses ;

2) les personnes se déclarant vulnérables ;

3) les personnes (vulnérables ou peu vulnérables) qui déclarent être exposées à un risque important d'avoir contracté le Covid-19 et d'être contagieuses.

En rendant impossible la distinction entre les utilisateurs des catégories 2 et 3, l'application encourage les utilisateurs de la catégorie 1 représentant la très grande majorité de la population à se tenir indistinctement à distance des utilisateurs des catégories 2 et 3 ; à éviter d'attraper la maladie à proximité d'une personne dont les risques de contagiosité sont élevés ; et à prendre le risque de contaminer, sans le savoir, une personne vulnérable.

En intégrant dans cette application les personnes se déclarant comme potentiellement contagieuses, l'application dont l'objectif est essentiellement de protéger les personnes vulnérables pourrait obtenir plus d'utilisateurs parmi les personnes non contagieuses et peu vulnérables, en téléchargeant l'application dans le but principal d'éviter les personnes contagieuses.

L'hypothèse probable d'utilisateurs se déclarant (sciemment ou inconsciemment) comme contagieux sans l'être réellement n'est paradoxalement pas un problème pour le bon fonctionnement de l'application en termes de dissuasion.

Cette approche ternaire, par sa simplicité et son caractère ludique, peut populariser son utilisation auprès des jeunes générations.

Une telle approche faciliterait la mise à l'échelle de l'application tout en renforçant l'approche «éthique dès la conception » en limitant la discrimination fondée sur la vulnérabilité ou la contagiosité.

7. CONCLUSION

Dans ce papier, nous avons tenté de discuter de l'utilité des applications de sécurité publique dans le processus de déconfinement. Tout d'abord, nous avons étudié et discuté les solutions existantes. A notre connaissance, la plupart de ces applications proposent de tracer les contacts afin d'évaluer le risque d'être confronté au COVID-19. Ensuite, nous avons proposé une idée qui suit une philosophie différente, axée sur la vie privée et l'identité auto-souveraine. Au lieu de rechercher les personnes touchées par le COVID-19, notre approche vise à contribuer à assurer la distanciation sociale, à protéger les personnes

vulnérables sans divulguer de données sensibles sur la santé. Nous avons développé une application open source qui respecte cette idée. Elle est basée sur une approche innovante et inclusive de l'évaluation de la vulnérabilité basée sur la théorie de la fonction de croyance et les techniques d'apprentissage automatique (*ML*). En outre, cette application met en œuvre un protocole de preuve à divulgation nulle de connaissance non-interactif pour garantir l'anonymat. Enfin, nous avons discuté des limites et des éventuelles évaluations futures de notre application. L'application uTakeCare pourrait être adaptée à l'avenir pour d'autres épidémies comme la grippe saisonnière, la dengue ou le paludisme par exemple ou plus généralement comme un outil contribuant à une citoyenneté anonyme et numérique.

POLITIQUE EN MATIÈRE DE CODE SOURCE

« uTakeCare » est une application open source. Nous encourageons la communauté scientifique à dupliquer et s'approprier le dépôt et à l'adapter à leurs propres fins. Nous encourageons la collaboration de la communauté open-source, et les demandes d'extraction seront prises en compte pour être incluses dans la branche principale du paquet. Veuillez consulter le site https://utakecare.fr pour les instructions.

RECONNAISSANCE

Nous tenons à remercier Koussay Dellagi (Institut Pasteur), pour les discussions à l'origine de "*uTakeCare"*.

RÉFÉRENCES

[1] M. Ienca, E. Vayena (2020). On the Responsible Use of Digital Data to Tackle the COVID-19 Pandemic. *Nature Medicine Volume 26. pp 463-464.* Mars 2020. DOI : https://doi.org/10.1038/s41591-020-0832-5

[2] D. Lee, J. Lee (2020). Testing on the Move South Korea's Rapid Response to the COVID-19 Pandemic. Transportation Research Interdisciplinary Perspectives. pp 1-33. April 2020. Advance online publication. https://doi.org/10.1016/j.trip.2020.100111

[3] CNIL (Commission nationale de l'informatique et des libertés), Délibération N° 2020-046 du 24 avril 2020 portant avis sur un projet d'application mobile dénommée « StopCovid », 2020 Apr 24. *Access :* https://www.mafr.fr/fr/article/deliberation-n-2020-046-du-24-avril-2020-portant-a/

[4] INRIA (Institut National de Recherche en Informatique et en Automatique) France (2020). Proximity Tracing Applications: *The misleading debate about centralised versus decentralised approaches*. pp 1-3. April 2020. *Access:* https://github.com/ROBERT-proximity-tracing/documents

[5] L. Kelion (2020). Coronavirus: Apple and Google Team Up to Contact Trace Covid-19. BBC News, 10 (04), April 2020 . https://www.bbc.com/news/technology-52246319

[6] S. Vaudenay (2020). Analysis of DP3T. April 2020. Access : https://github.com/DP-3T/documents

[7] H. Mayer (2020). zk-SNARK explained: Basic Principles. pp 1-8. December 2018. DOI : 10.13140/RG.2.2.20887.68007

[8] N. Chen, M. Zhou, X. Dong, J. Qu, F. Gong, Y. Han, Y. Qiu, J. Wang, Y. Liu, Y. Wei, J. Xia, T. Yu, X; Zhang, L. Zhang (2020). Epidemiological and Clinical Characteristics of 99 Cases of 2019. Novel Coronavirus Pneumonia in Wuhan, China: A Descriptive Study. *Lancet online Journal, Volume 395:10223. pp 507-513*, January 2020. DOI:10.1016/S0140-6736(20)30211-7

[9] ECDC experts : C. Adlhoch, N. Alberska, B/ Albiger, L. Alexakis, A. Baka, E. Broberg, S. Brusin, N. Bundle, M. Catchpole, ... eds (2020). Rapid risk assessment: Coronavirus Disease 2019 (COVID-19) pandemic: Increased Transmission in the EU/EEA and the UK – eighth update - 6 April 2020. *European Centre for Disease Prevention and Control report. pp 1-39*. April 2020.

[10] S. Wang, Z. Chen, Y. Lin, L. Lin, Q. Lin, S. Fang, Y. Shi, X. Zhuang, Y. Ye, T. Wang, H. Zhang, C. Shao (2020). Clinical Characteristics of COVID-19 in Fujian Province: a Multicenter Retrospective study. *Research Square Online Journal. pp 1-20*. April 2020. DOI:10.21203/rs.3.rs-21268/v1

[11] D. DeCaprio, J. A Gartner, T. Burgess, S. Kothari, S. Sayed, C. J. McCall and S. Sayed (2020). Building a COVID-19 Vulnerability Index. *arXiv preprint arXiv:2003.07347. pp. 1-9.* Mars 2020. DOI: 10.1101/2020.03.16.20036723

[12] L. Jia, K. Li, Y. Jiang, X. Guo, and T. Zhao (2020). Prediction and Analysis of Coronavirus Disease 2019. *arXiv:2003.05447 [q-bio.PE]. pp 1-19*. March 2020.

[13] R. Dandekar and G. Barbastathis (2020). Quantifying the Effect of Quarantine Control in Covid-19 Infectious spread using Machine Learning. *medRxiv Journal. Cold Spring Harbor Laboratory Press publisher. pp. 1-13*. April 2020. DOI. 10.1101/2020.04.03.20052084

[14] M. Pourhomayoun and M. Shakibi (2020). Predicting Mortality Risk in Patients with COVID-19 Using Artificial Intelligence to Help Medical Decision-Making. *medRxiv Journal. Cold Spring Harbor Laboratory Press publisher*. pp. 1-13. April 2020. DOI. 10.1101/2020.03.30.20047308

[15] AFM. Batista, JL. Miraglia,THR. Donato, ADP. Chiavegatto Filho (2020). COVID-19 Diagnosis Prediction in Emergency Care Patients: A Machine Learning Approach. *medRxiv Journal. Cold Spring Harbor Laboratory Press*. pp. 1-8. April 2020. DOI. 10.1101/2020.04.04.20052092

[16] S. Garg , L. Kim, M. Whitaker, A. O'Halloran, C. Cummings et al. (2020). Hospitalization Rates and Characteristics of Patients Hospitalized with Laboratory-Confirmed Coronavirus Disease 2019 — COVID-NET, 14 States. MMWR Morb Mortal Wkly Rep 2020;69:458–464. March 1–30, 2020. DOI: http://dx.doi.org/10.15585/mmwr.mm6915e3external icon.

[17] Jason Wang, Chun Y. Ng et Robert H. Brook. (2020). How Taiwan Used Big Data, Transparency and a Central Command to Protect Its People from Coronavirus. Journal of the American Medical Association. March 3, 2020. https://jamanetwork.com/journals/jama/fullarticle/2762689

[18] Han D., Dezert J., Yang Y. (2014) New Distance Measures of Evidence Based on Belief Intervals. *In: Cuzzolin F. (eds) Belief Functions: Theory and Applications. BELIEF 2014. Lecture Notes in Computer Science, Volume 8764. Springe. pp 432-44.* September 2014. DOI https://doi.org/10.1007/978-3-319-11191-9_47

[19] Smarandache, F., Dezert J. (2020). Applications and Advances of DSmT for Information Fusion. American Research Press, Rehoboth.*Volume 3. ISBN-10: 1599730731. pp 1-760.* June 2009.

[20] nCoV-2019 Data Working Group. Epidemiological Data from the nCoV-2019 Outbreak: Early Descriptions from Publicly Available *Data. Accessed on* 2020-04-20: http://virological.org/t/epidemiological-data-from-the-ncov-2019-outbreak-early-descriptions-from-publicly-available-data/337.

[21] I. Serrato. COVID19-patient-outcome-forecast-Neural-Network-vs-Logistic-Regression. *Publicly Available data on Github repository. Accessed on* 2020-04-21: https://github.com/IsaSerrato/COVID19-patient-outcome-forecast---Neural-Network-vs-Logistic-Regression.

[22] Sudalai Rajkumar. Day level information on COVID-19 affected cases. *Kaggle repository. Accessed on 2020-04-15*: https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset/data.

[23] Singapore Ministry of Health (2020). Ministerial statement on whole-of-government response to the 2019 novel coronavirus (2019-NCoV). Singapore: Singapore Ministry of health. 2020. https://www.moh.gov.sg/news-highlights/details/ministerial-statement-on-whole-of-government-response-to-the-2019-novel-coronavirus-(2019-ncov)

[24] C. Dubruelh (2020). Côte d'Ivoire: Une Application pour Limiter la Propagation du Covid-19. CIO magazine. Avril 2020. https://cio-mag.com/cote-divoire-une-application-pour-limiter-la-propagation-du-covid-19/

[25] Salla Gueye. "Covid-Trace": Une application 100% sénégalais pour tracer les cas contacts de coronavirus.. 2020-04-25. https://www.seneweb.com/news/Sante/laquo-covid-trace-raquo-une-application-_n_315795.html

[26] La Rédaction (2020). Document: Comment le Maroc va tracer les contaminés au Covid-19. Access 2020-04-13: https://ledesk.ma/2020/04/13/document-comment-le-maroc-va-tracer-les-contamines-au-covid-19/

[27] V. Kulkarni and P. Sinha (2013). Random Forest Classifiers: A Survey and Future Research Directions. *International Journal of Advanced Computing. Volume 36. pp 1144-1153*. January 2013.

[28] R. Raguram and J.M. Frahm and M. Pollefeys (2008). A Comparative Analysis of RANSAC Techniques Leading to Adaptive Real-Time Random Sample Consensus. Lect. *Note. Computer. Sciences. Volume 5303. pp 500-513.* April 2008. DOI: 10.1007/978-3-540-88688-4_37.

[29] L. Amour and S. Souihi (2020). New COVID-19 data set. *GitLab repository. Accessed on 2020-04-22:* https://gitlab.com/souihi/utakecare_prognosis/

[30] P. Smets (1993). Quantifying Beliefs by Belief Functions : An Axiomatic Justification. In Proceedings of the 13th International Joint Conference on Artificial Intelligence, *Proceedings of the 13th International Joint Conference on Artificial Intelligence - Volume 1. pp 598-605*, August 1993. DOI : 10.5555/1624025.1624109

[31] L. Amour, S. Souihi, A. Mellouk and S.M. Mushtaq (2019). Q2ABR: QoE-aware Adaptive Video Bit Rate Solution.. *In Journal of Network and Computer Applications. pages 1-13.* November 2019. DOI : https://doi.org/10.1002/dac.4204

[32] ITU News (2020). Ghana Launches COVID-19 Tracker App. ITU News. April 2020. *Accessed on: 16-04-2020 :* https://news.itu.int/ghana-launches-covid-19-tracker-app/

[33] Australian Government Department of Health (2020). COVIDSafe App. *Access on* 2020-05-08 https://www.health.gov.au/resources/apps-and-tools/covidsafe-app

[34] Gobierno Nacional de Colombia (2020). CoronApp. April 2020. Access on : 2020-04-19. https://coronaviruscolombia.gov.co

[35] District Health Information Software 2 (2020). COVID-19 Surveillance Digital Data Package. *Health Information Systems Program (HISP) at the University of Oslo (UiO). Mars 2020.* Access on : 2020-03-27. https://www.dhis2.org/covid-19

[36] J. Zaugg (2020). Singapour subit de Plein Fouet la Seconde Vague de Covid-19. *Le Temps Newsletter.* April 2020. Access on : 2020-04-29. https://www.letemps.ch/monde/singapour-subit-plein-fouet-seconde-vague-covid19

[37] J. Bay (2020). Automated Contact Tracing is not a Coronavirus Panacea. *Government Digital Services - GovTech Singapore.* April 2020. Access on : 2020-04-14 :https://blog.gds-gov.tech/automated-contact -tracing-is-not-a-coronavirus-panacea-57fb3ce61d98

[38] Smart Nation and Digital Government office (2020). Implementing SafeEntry and Safe Management practices. *Ministry of Health Singapore website.* April 2020. Access on : 2020-05-09: https://www.moh.gov.sg/news-highlights/details/implementing-safeentry-and-safe-management-practices

[39] S. Charma (2020). Aarogya Setu has 50 Million Users in 13 Days, Beats 'Pokémon GO' record. *Hindustan Times, New Delhi.* April 2020. Access on : 2020-04-15 *: https://www.hindustantimes.com/india-news/aarogya-setu-has-50-million-users-in-13-days-beats-pokemon-go-record/story-4Q25vLRuezSuzPA8jboLEL.html*

[40] D. C. D. Cruze (2020). Aarogya Setu Now Mandatory for Employees in India. Here's How to Register. Livemint. *National Informatics Centre. Government Ministry of Electronics and Information Technology in India.* May 2020. Access on :2020-05-02.

https://www.livemint.com/news/india/aarogya-setu-now-mandatory
-for-employees-in-india-here-s-how-to-register-11588408846545.html
[41] ZeroFOX Alpha Team.Iran (2020). Colombia and Italy Put Citizens at Risk with COVID-19 Government Mobile Apps. *ZeroFOX.* April 2020. Access on : 2020-04-06. https://www.zerofox.com/blog/covid-19-mobile-apps/
[42] S. Coble (2020). Vulnerabilities Detected in Government-sanctioned COVID-19. *App. Infosecurity-Magazin*e. April 2020. Access on : 2020-04-06. https://www.infosecurity-magazine.com/news/ vulnerabilities-covid19-app/
[43] J. P. Sharma (2020). Make Aarogya Setu Source Data Public For Greater Transparency, Scrutiny: Experts. *Outlook India.* May 2020. Acces on 2020-05-07 : https://www.outlookindia.com/website/story/
make-aarogya-setu-source-data-public-for-greater-transparency-scrutiny-experts/352236
[44] University of Oxford. Digital contact tracing can slow or even stop coronavirus transmission and ease us out of lockdown. *Nuffield Department of Medicine (NDM).* April 2020. Access on : .2020-04-16.
https://www.research.ox.ac.uk/Article/2020-04-16-digital-contact-tracing-can-slow-or-even-stop-coronavirus-transmission-and-ease-us-out-of-lockdown