Public Health Section

Machine Learning Models in Prediabetes Screening: A Systematic Review

AZMAWATI MOHAMMED NAWI[1], PUTERI SOFIA NADIRA MEGAT KAMARUDDIN[2], NOR RUMAIZAH MOHD NORDIN[3], SHARIFAH SAFFINAS SYED SOFFIAN[4], MAZNI BAHAROM[5]

(cc) BY-NC-ND

ABSTRACT

Introduction: The increasing prevalence of type 2 Diabetes Mellitus (DM) can be done from identifying those with prediabetes and offer early interventions by utilising prescreening diagnostic tools. Machine learning algorithms and big data mining approaches have been postulated for predictive disease modelling in hospital and clinical settings.

Aim: To outline the relative performance accuracies in predicting prediabetes conditions in different machine learning algorithms.

Materials and Methods: A systematic literature search was conducted at Universiti of Kebangsaan, Kuala Lumpur, Malaysia, based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) review protocol, and the research question was formulated based on the keywords of "Prediabetes" (Population), "Internet of Things" and "prediction model" (Intervention) and "screening" and "risk" (Outcome). International Prospective Register of Systematic Reviews (PROSPERO) registration (CRD42021264947) was done and databases were screened on 10th-24th June 2021 via Web of Science, Scopus, PubMed, Ovid and EBSCOhost. Inclusion criteria was English language prediction studies published between 2011-2021.

Review articles, editorials, proceedings, commentary articles and articles not focusing on prediabetes were excluded. The quality of the articles was ranked via the Prediction Model Risk of Bias Assessment Tool (PROBAST).

Results: A total of five articles that were published in 2014-2021 were included. The sample sizes ranged from 570 to 24,331 participants. Three studies (South Korea, United State of America (USA), Japan) suggested the applicability of the screening score prediction models for use in clinical settings related to personalised risk assessment and targeted interventions, with the predictors used being suitable for either the clinic or hospital. The simplicity of gender, age, Body Mass Index (BMI), blood pressure and waist circumference as predictors suggested that they can be utilised by the community.

Conclusion: This review highlights the fact that the heterogeneity of the population used and validation issues may affect generalisation. Future studies should address these concerns to guide advocacy among healthcare providers in clinical practice as well as in data and expertise sharing for developing and validating urgently needed prediabetic prediction models.

INTRODUCTION

Diabetes Mellitus is one of the world's most serious public health problems, causing premature death and imposing a heavy global disease burden. It is among the top ten causes of death, causing an estimated 1.5 million deaths globally in 2019 [1]. The overall global burden of diabetes has increased significantly and will continue to rise in the next few decades. There was an increment of 102.9% in the global incidence of diabetes from 11.3 million in 1990 to 22.9 million in 2017. Global Diabetes-Associated Disability Adjusted Life Years (DALYs) increased by 116.7% from 32.3 million in 1990 to 70.4 million in 2019 [2]. In 2021, an estimated 537 million adults had diabetes, and almost half (240 million) of them were undiagnosed [3]. The prevalence of diabetes is predicted to rise to 570.9 million in 2025 and to 693 million in 2045 [2,4]. If this particular concern remains undetected and undiagnosed, it will be left untreated. Furthermore, it can lead to potential complications such as heart disease, kidney disease, diabetic retinopathy and neuropathy. The complications of DM impose significant adverse health impacts and economic burdens on countries and their healthcare systems [5].

Prediabetes is a condition where blood glucose levels are higher than normal but below the diabetes threshold (Fasting Plasma Glucose (FPG) ≥7.8 and <11.1 mmol/L) [6]. It is an intermediate stage between DM and normal glucose tolerance and is defined either as impaired fasting blood glucose or Impaired Glucose Tolerance (IGT) [7]. It is regarded as a high risk condition, with a high likelihood of progressing to

diabetes [8]. Annually, approximately 5-10% of people with prediabetes develop diabetes; however, the conversion rates vary depending on the definition of prediabetes and the population characteristics [9,10]. In 2017, an estimated 7.7% of the world population (374 million) had IGT, and this number is expected to increase to 587 million (equal to 8.4%) in 2045 [4]. It has been established that lifestyle modifications and pharmaceutical interventions reduce the incidence of diabetes by an average of 20% relative risk reduction [11]. Thus, one strategy for addressing the increasing prevalence of type 2 DM is to identify those with prediabetes and offer such interventions.

Risk assessment tools can be designed to predict the likelihood of a particular health outcome based on a person's attributes and risk factors. By allowing screening to be focused on people at the highest risk, risk assessment tools aid optimisation of the resources necessary for identifying illnesses, which are typically limited [12]. Thus, risk assessment tools are useful for identifying people with prediabetes who may benefit from intervention, with many advocating them as the first stage in a screening programme [13]. Prescreening diagnostic tools have enabled clinicians to make better judgments and diagnose patients more quickly [14]. Any delay in disease detection might result in irreversible complications, such as blindness in diabetic retinopathy or end-stage renal failure in diabetic nephropathy [15]. Given the extensive screening and diagnosis processes required to protect the population from a variety of serious illnesses, certain healthcare sectors may be experiencing a shortage of diagnostic experts. As a result,

computer assisted technology can be used to aid in the prescreening process, resulting in improved diagnosis and prognosis.

Machine learning is an area of artificial intelligence research that tries to learn from past experiences and uses tools such as statistics, probabilistic and optimisation algorithms for classifying newly input data [16]. Previously, a powerful statistical analysis such as multivariate regression or correlation analysis was effective for constructing models by linearly combining the relevant variables [17]. However, the digitisation of medical records has resulted in a wealth of multidimensional data being stored in health databases. It represents a unique opportunity for advanced machine learning approaches to pattern recognition and prediction [18]. Unlike the traditional statistical method, machine learning methods use a wide range of parameters such as Boolean logic, absolute restriction, conditional probabilities, and unconventional optimisation methods for classification, nearly resembling that of a human being. Although most machine learning approaches draw concepts from statistics and probability, it has become a more powerful classification tool because it can generate a decision or inference from a dataset that conventional statistical techniques cannot [19].

Numerous machine learning techniques have been applied in clinical settings for the purpose of disease prediction, and have demonstrated a higher diagnostic accuracy than conventional methods [19]. Support Vector Machines (SVM), Artificial Neural Networks (ANN), Naïve Bayes algorithm and Random Forest (RF) are widely used machine learning approaches in disease risk prediction [20]. Machine learning algorithms and big data mining approaches have improved diabetes screening and prediction [21,22]. Given the increased applicability and effectiveness of machine learning algorithms for predictive disease modeling in hospital and clinical settings, we discovered little research that provides a complete evaluation of published publications that used machine learning algorithms for predicting prediabetes [23]. Therefore, this review is aimed at outlining the relative performance accuracies of different machine learning algorithms in predicting prediabetes conditions.

MATERIALS AND METHODS

The present systematic review was initiated at Universiti of Kebangsaan Malaysia, Kuala Lumpur, Malaysia and guided by the PRISMA review protocol [24]. A systematic literature search was conducted and registered under PROSPERO (CRD42021264947) and can be retrieved at https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42021264947 [25]. The PRISMA protocol aims to prompt researchers to source the right information with an accurate level of detail. Based on this protocol, we started the systematic literature review by formulating the appropriate research question. The systematic search consisted of three main sub-processes: identification, screening, inclusion.

Research question formulation: Here, the formulation of the research question was based on Patient/Population, Intervention, Comparison and Outcomes (PICO), a tool that aids authors in developing a suitable research question for a review. It is based on three main concepts: Population or problem, interest or intervention, and context or comparison and outcome [26]. Based on these concepts, the three main aspects included in the review were prediabetes (population), the Internet of Things and prediction models (intervention) and screening and risk (outcome), which guided the formulation of the main research objective.

Systematic Search Strategy

The three main processes in the systematic search strategy were identification, screening and inclusion.

Identification: The identification process involved searching for any synonyms, Medical Subject Heading (MeSH) terms, related terms and variation of the main keywords: prediabetes, Internet of Things, prediction model, screening and risk [Table/Fig-1]. This process provided greater coverage for finding related articles in the selected databases (Web of Science, Scopus, PubMed, Ovid, EBSCOhost) for the literature search, which was conducted within two weeks (10th-24th June 2021). Some of the distinct features of these databases were a large collection of literature and advanced search functions. We obtained 776 articles from the databases used; 28 duplicates were found and removed. The process returned 750 articles, and two articles were added from other sources, which were citations searching from the reference list of the initial included articles.

Screening: The 750 articles were screened with each database's sorting function. The article inclusion criteria were: journal article, written in English, published in 2011-2021, observational and interventional study with qualitative or quantitative data. We
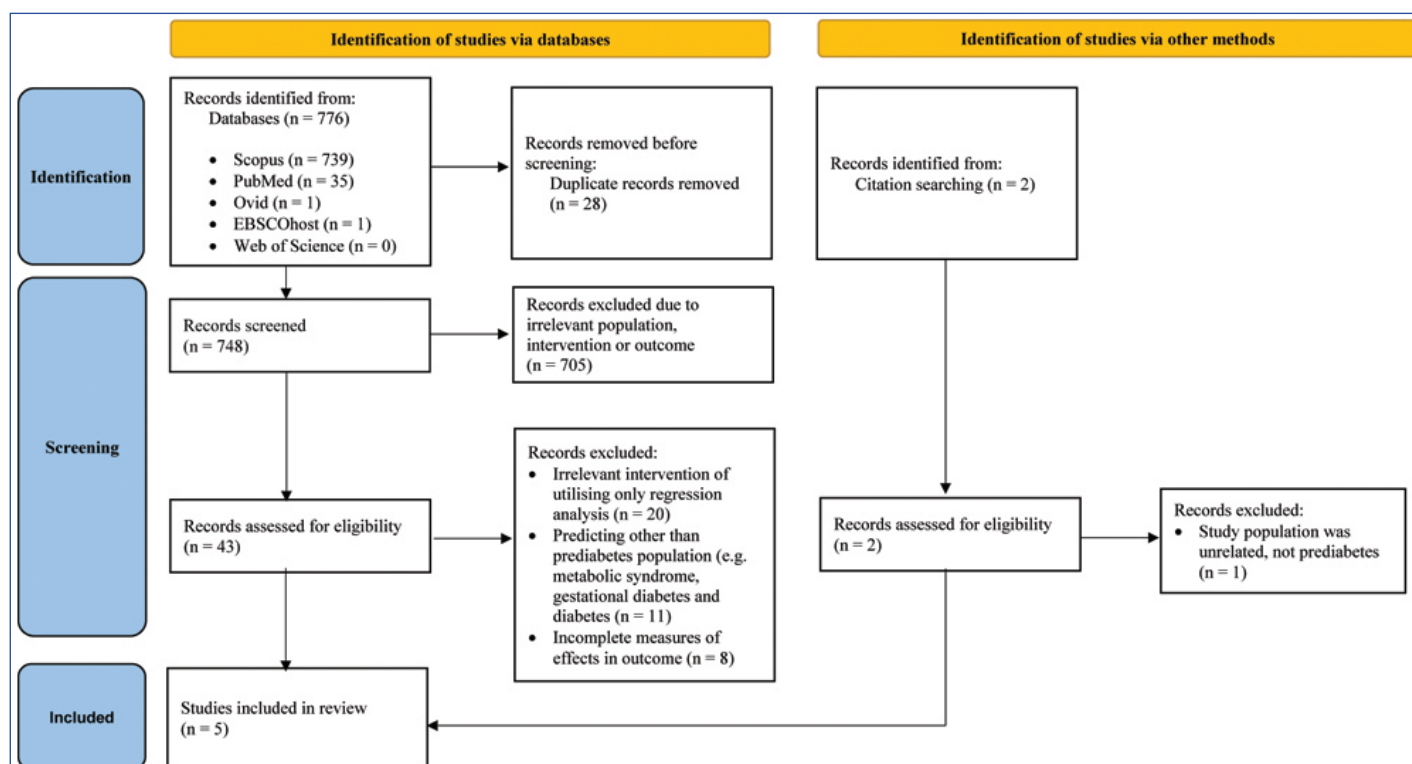
[Table/Fig-1]: PRISMA flow diagram [24].

excluded systematic reviews, comments or letters to the editor, abstracts from conferences, animal studies and in vivo or in vitro studies. Two teams of two or three authors independently screened the studies for inclusion. If there was disagreement at any stage, discussion leading to a consensus was made by a third author from the other team. We excluded 705 articles due to irrelevant population, intervention or outcome.

Eligibility for inclusion: The eligibility process was aimed at selecting the articles that fulfilled the study objective and was based on reading the article title and abstract. About 45 articles were manually sorted articles that satisfied the outcome of usage or development of machine learning models in prediabetes screening. Studies not related to the interest and intended outcome were excluded. Via this process, 36 articles were excluded based on irrelevant intervention, i.e., utilising only regression analysis; incomplete measures of effects on outcome; and predicting other than the prediabetes population, for example, metabolic syndrome, gestational diabetes and diabetes. In the final eligibility process, only five articles were included [Table/Fig-2] [27-31].

Data Extraction and Analysis

Thematic analysis was used in the present systematic review because it is considered in synthesising and integrating mixed research designs [32]. Thematic analysis is a descriptive analysis that allows data to be merged with other data analysis techniques [33]. The selected five articles were read in detail, especially the abstract, method, results and discussion sections. Then, the data were extracted based on whether the study was able to answer its own research questions, and the findings were simplified as tabulated in [Table/Fig-3] [27-31]. Only after these lengthy processes thematic analysis was done. To generate relevant themes, each author identified patterns of extracted data from the reviewed articles and grouped them before successfully categorising them into different themes of screening tool in regards to practicality, usability, generalisation, missing data and validation. The themes accuracy, usefulness and accurate data representation were re-reviewed. The developed themes were then submitted to a group of panel experts well versed in systematic reviews and in public health related research. The panel expert group subsequently agreed on the themes generated as being appropriate and accurate to the results of the review.

Quality Appraisal

The quality of the final list of studies was ranked according to the PROBAST to facilitate an objective assessment of the Risk Of Bias (ROB) and relevance of studies that established, validated or latest

Population	Intervention 1		Intervention 2	Outcome 1	Outcome 2
Prediabetes	Internet of things	Regression and classification	Prediction model*	Screening*	Risk
Asymptomatic diabetes	Artificial intelligence	Artificial narrow intelligence	Predicitive analytics	Detection	Possibility
Prediabetic state*	Machine learning	Artificial general intelligence	Predict future events	Screening test	Chance
Impaired glucose tolerance	Universal daemonisation	Artificial superintelligence	Analyse patterns	Early diagnosis	Probability
Impaired fasting glucose	Universal object interaction (UOI)	Neural network	Forecast future results	Diagnosis	Likelihood
Impaired glucose intolerance	Programmable onject interfaces (POI)	Development of 'thinking' computer systems	Predict future behaviour	Diagnose*	Prospect
Borderline diabetes	Transfurigated phase inversion	Machine like	Data-mining technology	Early detection	Uncertainty
Predevelopment diabetes	Internet-connected objects	Machine made	Forecast model*	Research	Hazard
	Prediction model	Machine minder	Classification model*	Check	Imminence
	Artificial neural network	Machine operator	Predicitve model*	Investigation	Unpredictability
	Decision tree	Machine language	Outliers model*	Analysis	
	Support vector machine	Machine instruction	Time series model	Assessment	
	Robotics	Neural network	Clustering model	Proof	
	Expert system*	Artificial insemination	ANOVA	Evaluation	
	Intelligent retrieval	Human language technology	Linear regression	Assessment	
	Knowledge engineering	Computer technology	Logistic regression	Assess	
	Natural language processing	Automatic learning	Ridge regression	Assay	
	Reactive machines	Learning machine	Random forest	Early detection	
	Limited memory	Information technology	Gradient boosted model	Health check*	
	Theory of mind	Semantic net	K-means	Medical check*	
	Self-awareness	Semantic network	Prophet	Medical check-up	
	Artificial superintelligence	Computer architecture	Prediction interval	Medical test*	
	Machine learning algorithms	Linear regression	Predictive inference	Health assessment	
	Supervised learning	Logistic regression		Health surveillance	
	Unsupervised learning	Linear discriminat analysis		Medical inspection	
	Reinforcement learning	Classification and regression times		Prevent	
		Naive bayes			
		K-nearest neighbours			
		Learning vector quantisation			
		Deep learning			
		Machine learning engineer			

[Table/Fig-2]: Keywords used for systematic searching.
*Truncation or wildcard symbol during keywords search

Authors, year, country	Database	Sampling frame	Sample size	Applicability	Prediabetes definition	Variable use as predictors/ Risk factors	Type of machine learning	Validation	Missing data	Outcome
Choi SB et al., 2014 South Korea [27]	Data from the Korean National Health and Nutrition Examination Survey (KNHANES), 2010	Adult South korean (age 51 to 54 years old)	4,685 subjects excluding subjects with diabetes	Clinical applications related to personalised risk assessment and targeted interventions	Fasting glucose 100-125 mg/dL (5.6-6.9 mmol/L) OR 2-hour OGTT measurement in 140-199 mg/dL (7.8-11.0 mmol/L) OR HbA1c 5.7-6.4% (39-46 mmol/L)	Age, gender, family history of diabetes, alcohol intake, BMI, waist circumference, FPG, systolic and diastolic blood pressure, HPT	Artificial Neural Network (ANN), Support Vector Machine (SVM)	Internally, 10 fold cross validation (KNHANES 2010) External validation: (data from KNHANES 2011)	Missing data for waist circumference, smoking status, alcohol intake, BMI, physical activity, family history of diabetes were excluded	Receiver operating characteristics (ROC) for accuracy Internal Set AUC=0.768 (ANN) 0.761 (SVM) External set AUC=0.729 (ANN) 0.731 (SVM)
Anderson JP et al., 2016 USA [28]	Electronic health record from Humedica database of US adult (20 integrated healthcare delivery network (IDN) in the US), 2007-2012	US adults belong to any integrated healthcare delivery network in the Humedica database	24,331 US adults without type 1 diabetes, enter the database with blood glucose measure in the low-risk range (normogly-caemic)	Clinical applications related to personalised risk assessment and targeted interventions	Fasting glucose ≥110 mg/dL OR two hour OGT ≥140 mg/dL OR random glucose ≥140 mg/dL OR HbA1c >5.7% *American Diabetes association guideline	Age, BMI, HDL, Triglycerides, ALT, CRP, body temperature	Reverse Engineering Forward Simulation (REFS), relying on bayesian scoring algorithm	Externally with the non IDN sets of population	Patients with missing values were not removed, instead were modeled as discrete with a missing category	Receiver operating characteristics (ROC) for accuracy AUC=0.72
Maeta K et al, 2018, Japan [29]	Comprehensive periodic medical examinations of Employee at the Center for Preventive Medicine, NTT Medical Center Tokyo, 2006-2017	Healthy office workers (age 40 to 60 years)	9,906 subjects without serious diabetes or advanced renal failure	Clinical applications related to personalised risk assessment and targeted interventions	Fasting glucose ≥110 mg/dL OR two hour OGT ≥140 mg/dL OR random glucose ≥140 mg/dL OR HbA1c >5.7%) *American Diabetes association guideline	OGTT (one hour Plasma glucose, one hour Immunoreactive insulin, two hour plasma glucose, two hour Immunoreactive insulin)	XGboost, relying on Gradient Tree Boosting (GTB)	Internally, five fold cross validation using test data subset	Missing data were excluded	ROC for accuracy (prediabetes or diabetes) AUC=0.75-0.78 *2 model of ML using complete OGTT as main variable *outperform Logistic Regression (LR) Model
Abbas M et al, 2021, Qatar [30]	Clinical, anthropo-metric, demographic data from qatar bio bank, data since 2012	Adult Chinese (age 18 to 86 years old)	7,268 participants (excluding: BMI <18.5 kg/m²)	Clinical and community settings, by primary healthcare physicians and by the general public	HbA1c: 39 mmol/L (5.7%)-47 mmol/L (6.4%) *American Diabetes association guidelines	Sex, age, BMI, blood pressure and waist circumference	Random Forest (RF), Gradient Boosting Machine (GBM), XgBoost (XGB) and Deep Learning (DL)	Internally, 10 fold cross validation using validated set data for all ML	All the variables had <10% missing values, which were imputed using the MICE package in R version 3.5.2	ROC for accuracy AUC = 0.81 (DL) 0.81 (GBM) 0.81 (XGB) 0.81 (RF) *ML did not outperform LR AUC=0.80
Li et al., 2021, China [31]	Data from Shuguang Hospital affiliated with Shanghai University of Traditional Chinese Medicine	Prediabetic population (age 57 to 68 years old)	570 prediabetic subjects	Hospital setting only (tongue images need to be collected by specially trained researchers)	Two hour postprandial blood glucose ≥7.8 mmol/L but <11.1 mmol/L OR fasting blood glucose ≥6.1 mmol/L but <7.0 mmol/L OR HbA1c ≥5.7% but <6.5%.	Tongue image data sets (deep feature, color and texture feature, fusion of features)	Genetic Algorithm-Extreme Gradient Boosting (GA-XGBT)	Internally, 10 fold cross validation using training set	No mention of missing data	ROC for accuracy AUC=0.93 (colour and texture) 0.816 (deep feature) 0.914 (fusion)

[Table/Fig-3]: Research findings related to machine learning prediction model for prediabetes [27-31].
OGT: O-linked; N: Acetylglucosamine (GlcNAc) transferase; OGTT: Oral glucose tolerance test; FPG: Fasting blood sugar; HPT: High pressure test; HDL: High density cholesterol; ALT: Alanine transamine; CRP: C-reactive protein; Independent distribution network; IDN: Integrated delivery network; ML: Machine learning; MICE: Multivariate imputation by chained equations; HbA1c: Glysated haemoglobin

prediction models for individualised predictions in a focused and transparent manner [34]. The PROBAST was recently developed by a steering group that considered existing ROB tools and reporting guidelines and was informed by a delphi procedure involving 38 experts, and refined through piloting [34]. In the present study, the teams extracted data from all included studies and assessed the ROB. If there was disagreement at any stage, discussion leading to a consensus was made by a third author from the other team. We performed qualitative analysis and appraisal of the included articles by extracting all relevant information using a predesigned standardised data extraction form. PROBAST is organised into the following four domains: participants, predictors, outcome, and analysis of different types of prediction studies, i.e., development or validation or both.

The PROBAST has a total of 20 key questions to help with the classification of ROB as low, high or unclear, and includes signalling questions to help make judgements. Signalling questions are rated as yes, probably yes, probably no, no or no information. Finally, overall judgement was made about the ROB and concerns regarding the applicability of the prediction model evaluation across all assessed domains [Table/Fig-4].

Authors, year, country	Type of prediction study?	Domain 1: Participants							Domain 2: Predictors						
		A. Risk of bias							A. Risk of bias						
		1.1 Were appropriate data sources used?	1.2 Were all inclusions and exclusions of participants appropriate?	Risk of bias introduced by selection of participants	Rationale of bias rating:	B. Applicability	Rationale of applicability rating:	2.1 Were predictors defined and assessed in a similar way for all participants?	2.2 Were predictor assessments made without knowledge of outcome data?	2.3 Are all predictors available at the time the model is intended to be used?	Risk of bias introduced by predictors or their assessment	Rationale of bias rating:	B. Applicability	Rationale of applicability rating:
Choi et al, 2014, South Korea [27]	Development & validation	Yes	Yes	Low	-	Low	-	Yes	Yes	Yes	Low	-	Low	-
Anderson JP et al., 2016, USA [28]	Development only	Probably No	Probably No	Low	Big data: 2,4331 with external validation	Low	Variables used are established risk factors for diabetes	Yes	Yes	Yes	Low	Realtime data were being used. missing data were treated as missing category resembling the real time situation. data obtained from registry hence no knowledge on the outcome	Low	Clear definition of predictors at the beginning of the model. cut off level by blood measurement made clear for each predictor
Maeta K et al., 2018, Japan [29]	Development only	Yes	Yes	Low	Large data set (20,458) with internal five fold cross validation	Low	-	Yes	Yes	Yes	Low	Predictors used in this model is well established and well defined	Low	-
Abbas M et a.l, 2021, Qatar [30]	Development only	Yes	Yes	Low	Large data set, with ten fold cross validation	Low	Variables used are established risk factors for diabetes	Yes	Yes	Yes	Low	Predictors used well defined/ established	Low	Clear definitions/ established with references
Li et al, 2021., China [31]	Development only	Yes	Probably Yes	Low	-	Low	-	Yes	Yes	Yes	Low	-	Low	-

Authors, year, country	Domain 3: Outcome									
	A. Risk of bias									
	3.1 Was the outcome determined appropriately?	3.2 Was a pre-specified or standard outcome definition used?	3.3 Were predictors excluded from the outcome definition?	3.4 Was the outcome defined and determined in a similar way for all participants?	3.5 Was the outcome determined without knowledge of predictor information?	3.6 Was the time interval between predictor assessment and outcome determination appropriate?	Risk of bias introduced by the outcome or its determination	Rationale of bias rating	B. Applicability	Rationale of applicability rating
Choi SB et al., 2014, South Korea [27]	Yes	Yes	Yes	Yes	Yes	No information	Low	-	Low	-
Anderson JP et al., 2016, USA [28]	Yes	Yes	Yes	Yes	Yes	Yes	Low	Distinct definition between the outcome and the predictors	Low	
Maeta K et al., 2018, Japan [29]	Yes	Yes	Yes	Yes	Yes	Yes	Low	Distinct definition between the outcome and the predictors	Low	-
Abbas M et al., 2021, Qatar [30]	Yes	Yes	Yes	Yes	Yes	Yes	Low	Distinct definition between the outcome and the predictors	Low	
Li et al, 2021, China [31]	Yes	Yes	Yes	Yes	Yes	Yes	Low	-	Low	-

Authors, year, country	Domain 4: Analysis — A. Risk of bias									Risk of bias introduced by the analysis	Rationale of bias rating:	Judgement	
	4.1 Were there a reasonable number of participants with the outcome?	4.2 Were continuous and categorical predictors handled appropriately?	4.3 Were all enrolled participants included in the analysis?	4.4 Were participants with missing data handled appropriately?	4.5 Was selection of predictors based on univariable analysis avoided?	4.6 Were complexities in the data (e.g., censoring, competing risks, sampling of controls) accounted for appropriately?	4.7 Were relevant model performance measures evaluated appropriately?	4.8 Were model overfitting and optimism in model performance accounted for?	4.9 Do predictors and their assigned weights in the final model correspond to the results from multivariable analysis?			Risk overall judgement of risk of bias	Concern overall judgement of applicability
Choi SB et al., 2014., South Korea [27]	Yes	Yes	Yes	Yes	No information	No information	Yes	Yes	No information	Low	To avoid overfitting, the prediction models were internally validated using cross validation.	Low	Low
Anderson JP et al, 2016, USA [28]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Low	Clear explanations given for all the domain analysed, supplemented with Kaplan Meier analysis and cox regression; overfitting of the model is considered	Low	Low
Maeta K et al., 2018, Japan [29]	Yes	Yes	Yes	Yes	No information	Yes	Yes	Yes	Yes	Low	To avoid overfitting, the prediction models were internally validated using cross validation.	Low	Low
Abbas M et al., 2021, Qatar [30]	Yes	Yes	Yes	Yes	No information	No information	Yes	Yes	Yes	Low	To avoid overfitting, the prediction models were internally validated using cross-validation.	Low	Low
Li et al, 2021, China [31]	Probably Yes	Yes	Yes	No information	No information	No information	Yes	Yes	No information	Low	The experimental data was inadequate Learning rate is step size shrinkage used in update to prevent overfitting 10 fold cross validation to evaluate the performance of XGBT models which were inputted different types of features.	Low	Low

[Table/Fig-4]: Screening for bias with PROBAST (Prediction Model Study Risk Of Bias Assessment Tool) [27-31].

RESULTS

Five articles were included in this review and had been published in 2014-2021. The sample sizes of the studies ranged from 570 to 24,331 participants. Four studies were from Asia: South Korea, Japan, Qatar and China. Only one study was from North America, i.e., the USA. The types of machine learning involved were Artificial Neural Network (ANN), Support Vector Machine (SVM), Reverse Engineering Forward Simulation (REFS), XGboost (XGB), RF, Gradient Boosting Machine (GBM), Deep Learning (DL) and GA-XGBT. All studies involved the development of prediction models for prediabetes screening. Three studies used the prediabetes definition based on American Diabetes Association guidelines, while the other two studies used different sets of definitions pertaining to their own country (South Korea and China). The differing units used and ranges resulted in slightly lower thresholds for FPG. Using PROBAST, all five studies had low ROB.

Study population and databases: The largest data set was obtained retrospectively for 2007-2012 from the electronic health records of the Humedica database on US adults, whereby 24,331 adults without type 1 diabetes were entered in the database with blood glucose in the low risk range (normoglycaemic) [28]. This was followed by the study from Japan, which analysed data from comprehensive medical check-ups (2006-2017) involving 9,906 healthy office workers without serious diabetes or advanced renal failure and who were aged 40-60 years [29]. The study from Qatar reported the clinical, anthropometric and demographic data of 7,386 people aged between 18 to 86 years from the Qatar Biobank, which has been collecting data from the general population since 2012, from which participants with Body Mass Index (BMI) <18.5 kg/m^2 were excluded [30]. The South Korean study used data from the 2010 Korean National Health and Nutrition Examination Survey (KNHANES)

involving 4,685 participants, excluding participants with diabetes, and involved South Korean adults aged 51-54 years [27]. Lastly, the study from China used 2011-2019 data from Shuguang hospital affiliated with Shanghai university of traditional chinese medicine, comprising 570 prediabetic participants aged 57-68 years [31].

Validation and missing data: Three studies performed internal validation using the same database, and it was either ten fold cross-validation or five fold cross validation [29-31]. The study from the USA performed external validation using datasets that were not from the Humedica database [28]. The South Korean study performed both internal (10 fold cross validation) and external validation using 2011 KNHANES data [27]. Regarding management of missing data, two studies excluded missing data [27,29] and two studies used imputation [28,30]. The study from China did not mention missing data [31].

ROC accuracy and applicability of machine learning in different settings: Three studies suggested the applicability of the screening score prediction models for use in clinical applications related to personalised risk assessment and targeted interventions [27-29]. The South Korean study used age, gender, family history of diabetes, alcohol intake, BMI, waist circumference, FPG and systolic and diastolic blood pressure as predictors in the ANN and SVM models, which showed Areas Under the Curve (AUC) of 0.768 and 0.761, respectively [27]. Anderson JP et al., reported an AUC of 0.72 for its REFS, which used age, BMI, HDL, triglycerides, alanine transaminase, CRP and body temperature as predictors [28]. In 2018, using a complete oral glucose tolerance test profile that consisted of one-hour plasma glucose, one-hour immunoreactive insulin, two-hour plasma glucose and two-hour immunoreactive insulin, reported AUC of 0.75-0.78 for the two XGboost models, which outperformed the logistic regression model [29]. Given the predictors used, these three machine learning algorithms are suitable for clinic and hospital settings.

The study from China used tongue image data sets, which consisted of deep feature, colour and texture feature and fusion of features. The authors reported that the GA_XGBT had an AUC of 0.93 (colour and texture feature), 0.816 (deep feature) and 0.914 (fusion of features) for predicting prediabetes [31]. Using tongue image datasets was only applicable to clinical application in hospital settings, due to the fact that tongue images need to be collected by specially trained researchers using specialised machines [31].

Gender, age, BMI, blood pressure and waist circumference were used as predictors in DL, GBM, XGB and RF, and AUC of 0.81 were observed from all four approaches. Moreover, the study from Qatar reported that these four machine learning approaches did not outperform the logistic regression model [30]. Due to their simplicity, the authors used gender, age, BMI, blood pressure and waist circumference as predictors, suggesting that other than the clinical setting, they can also be utilised by the community (general public) [16].

DISCUSSION

To the best of our knowledge, this is the first systematic review of the machine learning model approach used for screening people with prediabetes. The inclusion criteria were established in order to identify individuals who would benefit from interventions aimed at early detection and prevention of DM.

The machine learning based prediabetes risk score model provides an alternative screening tool that is inexpensive and simple to administer to people who appear to be healthy in the general population [35,36]. Compared to the traditional screening method of IGT testing, the prediction model incorporates other modifiable risk factors such as body temperature, smoking habit, BMI, tongue image information, blood pressure and waist

circumference, which does not require any additional procedure and is non invasive [28,30,31,35]. The implication is reflected in the convenience and practicality for further increasing early detection strategies in population-based settings.

Despite the high accuracy of the prediabetes prediction models identified in the present review, their usability in the clinical field is an important issue to be considered. The application of the prediction models at tertiary healthcare centres requires the attention of clinical experts as in the case of tongue feature [31], imposing additional resources and cost for screening purposes. Not withstanding, implementing an artificial intelligence model with standardised equipment for collecting and interpreting information may not be cost-effective in the overwhelmed hospital setting. In contrast, routine clinical measures are preferred and more practicable for aiding mass screening in the population.

With the expansion of electronic health records, more robust and advanced computational approaches such as machine learning have become the focus of disease prediction research [20,37]. The supervised machine learning model considers complex nonlinear relationships between dependent and independent variables of multiple data types [38]. The unique characteristic of an ANN includes a black box at the centre of the decision-making processes. However, it has been argued that the same machine learning model may produce results with varying degrees of accuracy for the same dataset based on the selection of different underlying parameters [35,39]. The high precision generates exclusive terms for specific populations, thus posing challenges to generalisation across different settings.

In general, the characteristics of the data used for developing the prediction models were well described; however, a few models sampled specific databases according to employment status and tertiary healthcare centres rather than population-based sampling [29,31]. Recruitment through specific databases is likely to result in a non representative population and hence should be avoided [40,41]. Future use of such prediction models should be approached cautiously, as the outcome will be accurate only for screening in a similar group of people.

Furthermore, it is imperative that model development studies clearly explain the treatment of missing data [38,39,42]. The majority of the prediction model developers analysed in the present review opted to exclude missing data rather than use imputation to assign plausible values to fit in [29,31,43]. Multiple imputation produces more valid results and better discrimination compared to excluding missing data upon integration into the model because it accounts for the average values of the parameters used within a model, hence reading through the trends [20,44]. By including the missing data, a prediction model justifies the true real world population characteristics.

Validation of the prediction model is crucial for evaluating the discrimination ability and calibration to support the stability of the suggested model [20]. External validation, regarded as the gold standard, should be performed prior to considering a prediction tool for use in the real world setting [37,41]. However, only two of the studies included in the present review reported on external validation. Whereas the prediction model ensembles performed reasonably well in predicting prediabetes, with AUC ranging between 0.72 and 0.93, critical observations should be made during replication in other study populations for potential bias.

Evaluating the application of the prediction model in clinical practice is vital before proceeding with advocacy activities [40,45]. All of the articles included in this review discussed the subsequent impact on healthcare practises, highlighting the importance of a greater emphasis on model use prior to development. Based on personalised risk assessment, the healthcare provider can plan more targeted interventions tailored to the person's need [28-30].

With the growing nature of electronic health information data-keeping, prediction models using machine learning provide better usability for the healthcare provider to stratify patients according to different risk factors.

The prediction models for prediabetes reviewed here incorporate the computational approach, utilising big data for model development. Big data enables greater insight into information on the real world population, hence the representativeness. The factor variables used in the model development accurately measure a specific individual characteristic, which is based on routine clinical parameters used in healthcare settings.

Limitation(s)

There are limited studies on the development of prediction models for prediabetes and machine learning to date, thus, it is difficult to assess the superiority of one model over another. In the present review, prediction models derived from certain populations, particularly high-income countries, may not be applicable to other populations with different regional settings due to the distinctive genetic makeups and socio-economic backgrounds.

CONCLUSION(S)

Prediction model studies on prediabetes are available and appear to show good accuracy outcomes. However, this review highlights the fact that the heterogeneity of the population used and validation issues may affect generalisation. Future studies should address these concerns to guide advocacy among healthcare providers. While the clinical data measured vary widely between prediction studies, a comparison is only possible when a common dataset benchmark is established. Therefore, there is an urgent need for data and expertise sharing for developing and validating prediabetic prediction models.

Author contributions: All authors contributed to the design and implementation of the research, analysis of the results and writing of the manuscript.

REFERENCES

[1] World Health Organization Dashboard available from https://www.who.int/news-room/fact-sheets/detail/diabetes accessed on 18th September 2021.
[2] Lin X, Xu Y, Pan X, Xu J, Ding Y, Sun X, et al. Global, regional, and national burden and trend of diabetes in 195 countries and territories: An analysis from 1990 to 2025. Scientific Reports. 2020;10(14790). Doi: https://doi.org/10.1038/s41598-020-71908-9.
[3] International Diabetes Federation. Diabetes Facts & Figures. 2021. Available from: https://idf.org/aboutdiabetes/what-is-diabetes/facts-figures.html#:#:~:text=The%20IDF%20Diabetes%20Atlas%20Tenth,and%20783%20million%20by%202045. Accessed on 20th October 2021.
[4] Cho NH, Shaw JE, Karuranga S, Huang Y, Fernandes JDdR, Ohlrogge AW, et al. IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045. Diabetes Research and Clinical Practice. 2018;138:271-81. Doi: https://doi.org/10.1016/j.diabres.2018.02.023.
[5] Meetoo D. Diabetes: Complications and the economic burden. British Journal of Healthcare Management. 2014;20(2). Doi: https://doi.org/10.12968/bjhc.2014.20.2.60.
[6] Bansal N. Prediabetes diagnosis and treatment: A review. World Journal of Diabetes. 2015;6(2):296-303. Doi: https://doi.org/10.4239/wjd.v6.i2.296.
[7] American Diabetes Association Dashboard. Available from: https://www.diabetes.org/a1c/diagnosis. Accessed on 20th October 2021.
[8] Tabák AG, Herder C, Rathmann W, Brunner EJ, Kivimäki M. Prediabetes: A high-risk state for diabetes development. Lancet. 2012;379(9833):2279-90. Doi: https://doi.org/10.1016/s0140-6736(12)60283-9.
[9] Forouhi NG, Luan J, Hennings S, Wareham NJ. Incidence of Type 2 diabetes in England and its association with baseline impaired fasting glucose: The Ely study 1990-2000. Diabetic Medicine. 2007;24(2):200-07. Doi: https://doi.org/10.1111/j.1464-5491.2007.02068.x.
[10] Nathan DM, Davidson MB, DeFronzo RA, Heine RJ, Henry RR, Pratley R, et al. Impaired fasting glucose and impaired glucose tolerance: Implications for care. Diab Care. 2007;30(3):753-59. Doi: https://doi.org/10.2337/dc07-9920.
[11] Sheng Z, Cao JY, Pang YC, Xu HC, Chen JW, Yuan JH, et al. Effects of lifestyle modification and anti-diabetic medicine on prediabetes progress: A systematic review and meta-analysis. Frontiers in Endocrinology. 2019;10:01-15. Doi: https://doi.org/10.3389/fendo.2019.00455.
[12] Khunti K, Gillies CL, Taub NA, Mostafa SA, Hiles SL, Abrams KR, et al. A comparison of cost per case detected of screening strategies for Type 2 diabetes and impaired glucose regulation: modelling study. Diabetes Research and Clinical Practice. 2012;97(3):505-13. Doi: https://doi.org/10.1016/j.diabres.2012.03.009.

[13] Gray LJ, Willis A, Webb D, Davies MJ, Khunti K. Screening for diabetes and prediabetes. In: Bonora E, DeFronzo R, editors. Diabetes epidemiology, genetics, pathogenesis, diagnosis, prevention, and treatment. Cham: Springer International Publishing; 2018. Pp. 1-33.
[14] Kumar U. Applications of machine learning in disease prescreening. In: Edoh T, Pawar P, Mohammad S, editors. Prescreening systems for early disease prediction, detection, and prevention: IGI Global; 2019. Pp. 278-320.
[15] Roche MM, Wang PP. Factors associated with a diabetes diagnosis and late diabetes diagnosis for males and females. Journal of Clinical and Translational Endocrinology. 2014;1(3):77-84. Doi: https://doi.org/10.1016/j.jcte.2014.07.002.
[16] Mitchell T. Machine Learning. 1st ed. United States of America: McGraw-Hill Education. 1997:344-50.
[17] Heltberg A, Andersen JS, Sandholdt H, Siersma V, Kragstrup J, Ellervik C. Predictors of Undiagnosed prevalent type 2 diabetes- The Danish general suburban population study. Primary Care Diabetes. 2018;12(1):13-22. Doi: https://doi.org/10.1016/j.pcd.2017.08.005.
[18] Prabhu LAJ, Sengan S, Kamalam GK, Vellingiri J, Gopal J, Velayutha P, et al. Medical information retrieval systems for e-Health care records using fuzzy based machine learning model. Microprocessors and Microsystems. 2020;2020:103344. Doi: https://doi.org/10.1016/j.micpro.2020.103344.
[19] Kumar U. Applications of Machine Learning in Disease Prescreening. 2018. Doi: 10.4018/978-1-5225-7131-5.ch010.
[20] Uddin S, Khan A, Hossain ME, Moni MA. Comparing different supervised machine learning algorithms for disease prediction. BioMed Central Medical Informatics and Decision Making. 2019;19:01-16. Doi: https://doi.org/10.1186/s12911-019-1004-8.
[21] Kavakiotis I, Tsave O, Salifoglou A, Maglaveras N, Vlahavas I, Chouvarda I. Machine learning and data mining methods in diabetes research. Computational and Structural Biotechnology Journal. 2017;15:104-16. Doi: https://doi.org/10.1016/j.csbj.2016.12.005.
[22] Marcos Lopez DP. Advances in Financial Machine Learning. 1st ed. Hoboken, New Jersey: John Wiley & Sons, Inc. 2018:246-55.
[23] Cahn A, Shoshan A, Sagiv T, Yesharim R, Goshen R, Shalev V, et al. Prediction of progression from pre-diabetes to diabetes: Development and validation of a machine learning model. Diabetes/Metabolism Research and Reviews. 2019;36(2):e3252. Doi: https://doi.org/10.1002/dmrr.3252.
[24] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. British Medical Journal. 2021;372:n71. Doi: https://doi.org/10.1136/bmj.n71.
[25] Mohd Nordin NR, Mohamed Nawi A, Kamaruddin PSNM, Baharom M, Soffian SSS. Machine learning models in pre-diabetes screening: A systematic review. PROSPERO 2021 [Available from: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42021264947.
[26] Pollock A, Berge E. How to do a systematic review. International Journal of Stroke. 2018;13(2):138-56. Doi: https://doi.org/10.1177/1747493017743796.
[27] Choi SB, Kim WJ, Yoo TK, Park JS, Chung JW, Lee Yh, et al. Screening for prediabetes using machine learning models. Computational and Mathematical Methods in Medicine. 2014;2014:618976. Doi: https://doi.org/10.1155/2014/618976.
[28] Anderson JP, Parikh JR, Shenfeld DK, Ivanov V, Marks C, Church BW, et al. Reverse engineering and evaluation of prediction models for progression to type 2 diabetes: An application of machine learning using electronic health records. Journal of Diabetes Science and Technology. 2016;10(1):06-18. Doi: https://doi.org/10.1177/1932296815620200.
[29] Maeta K, Nishiyama Y, Fujibayashi K, Gunji T, Sasabe N, Iijima K, et al. Prediction of glucose metabolism disorder risk using a machine learning algorithm: Pilot study. Journal of Medical Internet Research Diabetes. 2018;3(4):e10212. Doi: https://doi.org/10.2196/10212.
[30] Abbas M, Mall R, Errafii K, Lattab A, Ullah E, Bensmail H, et al. Simple risk score to screen for prediabetes: A cross-sectional study from the Qatar Biobank cohort. Journal of Diabetes Investigation. 2021;12(6):988-97. Doi: https://doi.org/10.1111/jdi.13445.
[31] Li J, Yuan P, Hu X, Huang J, Cui L, Cui J, et al. A tongue features fusion approach to predicting prediabetes and diabetes with machine learning. Journal of Biomedical Informatics. 2021;115. Doi: https://doi.org/10.1016/j.jbi.2021.103693.
[32] Flemming K, Booth A, Garside R, Tunçalp Ö, Noyes J. Qualitative evidence synthesis for complex interventions and guideline development: Clarification of the purpose, designs and relevant methods. British Medical Journal Global Health. 2019;4(Suppl 1):e000882. Doi: https://doi.org/10.1136/bmjgh-2018-000882.
[33] Vaismoradi M, Turunen H, Bondas T. Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. Nursing & Health Sciences. 2013;15(3):398-405. Doi: https://doi.org/10.1111/nhs.12048.
[34] Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A tool to assess the risk of bias and applicability of prediction model studies. Annals of Internal Medicine. 2019;170:51-58. Doi: https://doi.org/10.7326/M18-1376.
[35] Quan J, Ng CS, Kwok HHY, Zhang A, Yuen YH, Choi CH, et al. Development and validation of the CHIME simulation model to assess lifetime health outcomes of prediabetes and type 2 diabetes in Chinese populations: A modeling study. PLOS Medicine. 2021;18(6):e1003692. Doi: https://doi.org/10.1371/journal.pmed.1003692.
[36] Liberty IA, Kodim N. Assess prediabetes risk, as a golden period for prevention of diabetes. Asian Journal of Pharmaceutical and Clinical Research. 2017;10(6):349-53. Doi: https://doi.org/10.22159/ajpcr.2017.v10i6.18215.

[37] Wynants L, Calster BV, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of COVID-19: Systematic review and critical appraisal. British Medical Journal. 2020:369. Doi: https://doi.org/10.1136/bmj.m1328.

[38] Gurusamy V, Kannan S, Nandhini K. The real time big data processing framework advantages and limitations. International Journal of Computer Sciences and Engineering. 2017;5(12):305-12. Doi: https://doi.org/10.26438/ijcse/v5i12.305312.

[39] Barber SR, Davies MJ, Khunti K, Gray LJ. Risk assessment tools for detecting those with pre-diabetes: A systematic review. Diabetes Research and Clinical Practice. 2014;105(1):01-13. Doi: https://doi.org/10.1016/j.diabres.2014.03.007.

[40] Hong D, Fort D, Shi L, Price-Haywood EG. Electronic medical record risk modeling of cardiovascular outcomes among patients with type 2 diabetes. Diabetes Therapy. 2021. Doi: https://doi.org/10.1007/s13300-021-01096-w.

[41] Xue M, Su Y, Li C, Wang S, Yao H. Identification of potential type ii diabetes in a large-scale chinese population using a systematic machine learning framework. Journal of Diabetes Research. 2020;2020:6873891. Doi: https://doi.org/10.1155/2020/6873891.

[42] Choi SB, Lee W, Yoon JH, Won JU, Kim DW. Ten-year prediction of suicide death using Cox regression and machine learning in a nationwide retrospective cohort study in South Korea. Journal of Affective Disorders. 2018;231:08-14. Doi: https://doi.org/10.1016/j.jad.2018.01.019.

[43] Chung JW, Kim WJ, Choi SB, Park JS, Kim DW. Screening for pre-diabetes using support vector machine model. Computational and Mathematical Methods in Medicine. 2014:2471-75. Doi: https://doi.org/10.1109/EMBC.2014.6944123.

[44] Fujiati II, Damanik HA, Bachtiar A, Nurdin AA, Ward P. Development and validation of prediabetes risk score for predicting prediabetes among Indonesian adults in primary care: Cross-sectional diagnostic study. Interventional Medicine and Applied Science. 2017;9(2):76-85. Doi: https://doi.org/10.1556/1646.9.2017.18.

[45] Man B, Schwartz A, Pugach O, Xia Y, Gerber B. A clinical diabetes risk prediction model for prediabetic women with prior gestational diabetes. PLOS One. 2021;16(6):e0252501. Doi: https://doi.org/10.1371/journal.pone.0252501.

PARTICULARS OF CONTRIBUTORS:
1. Associate Professor, Department of Community Health, Faculty of Medicine, Universiti Kebangsaan Malaysia, Cheras, Wilayah Persekutuan Kuala Lumpur, Malaysia.
2. Postgraduate, Department of Community Health, Faculty of Medicine, Universiti Kebangsaan Malaysia, Cheras, Wilayah Persekutuan Kuala Lumpur, Malaysia.
3. Postgraduate, Department of Community Health, Faculty of Medicine, Universiti Kebangsaan Malaysia, Cheras, Wilayah Persekutuan Kuala Lumpur, Malaysia.
4. Postgraduate, Department of Community Health, Faculty of Medicine, Universiti Kebangsaan Malaysia, Cheras, Wilayah Persekutuan Kuala Lumpur, Malaysia.
5. Postgraduate, Department of Community Health, Faculty of Medicine, Universiti Kebangsaan Malaysia, Cheras, Wilayah Persekutuan Kuala Lumpur, Malaysia.

NAME, ADDRESS, E-MAIL ID OF THE CORRESPONDING AUTHOR:
Azmawati Mohammed Nawi,
Associate Professor, Department of Community Health, Faculty of Medicine,
Cheras, Wilayah Persekutuan Kuala Lumpur, Malaysia.
E-mail: azmawati@ppukm.ukm.edu.my

PLAGIARISM CHECKING METHODS: [Jain H et al.]
• Plagiarism X-checker: Nov 22, 2021
• Manual Googling: Feb 22, 2022
• iThenticate Software: Mar 22, 2022 (18%)

ETYMOLOGY: Author Origin

AUTHOR DECLARATION:
• Financial or Other Competing Interests: None
• Was Ethics Committee Approval obtained for this study?   NA
• Was informed consent obtained from the subjects involved in the study?   NA
• For any images presented appropriate consent has been obtained from the subjects.   NA

Date of Submission: **Nov 23, 2021**
Date of Peer Review: **Jan 08, 2022**
Date of Acceptance: **Feb 23, 2022**
Date of Publishing: **May 01, 2022**