Systematic review of diabetic foot ulcer classification models using artificial intelligence and machine learning techniques

Manuel Alberto Silva, Emma J Hamilton, David A Russell, Fran Game, Sheila C Wang, Sofia Baptista, Matilde Monteiro-Soares

Submitted to: Journal of Medical Internet Research
on: November 28, 2024

Table of Contents

Systematic review of diabetic foot ulcer classification models using artificial intelligence and machine learning techniques

Manuel Alberto Silva[1, 2] MD; Emma J Hamilton[3, 4] MD; David A Russell[5, 6] MD, PhD; Fran Game[7] MD, PhD; Sheila C Wang[8] MD, PhD; Sofia Baptista[9, 10, 11] MD, PhD; Matilde Monteiro-Soares[9, 10, 12, 13] PhD

[1]USF Sanus Carandá, ULS Braga Braga PT
[2]Faculty of Medicine, University of Porto Porto PT
[3]Department of Endocrinology and Diabetes, Fiona Stanley Hospital Murdoch AU
[4]University of Western Australia, Medical School, Fiona Stanley Hospital Murdoch AU
[5]Leeds Institute of Clinical Trials Research, University of Leeds Leeds GB
[6]Leeds Vascular Institute, Leeds Teaching Hospitals NHS Trust Leeds GB
[7]University Hospitals of Derby and Burton NHS Foundation Trust Derby GB
[8]Division of Dermatology, Department of Medicine, University of Toronto Toronto CA
[9]MEDCIDS - Departamento de Medicina da Comunidade Informação e Decisão em Saúde; Faculty of Medicine, University of Porto Porto PT
[10]CINTESIS@RISE - Center for Health Technology and Services Research; Faculty of Medicine, University of Porto Porto PT
[11]CUF Porto Hospital Porto PT
[12]Portuguese Red Cross Health School Lisbon Lisbon PT
[13]Cross I&D Lisbon PT

Corresponding Author:
Manuel Alberto Silva MD
USF Sanus Carandá, ULS Braga
Praça Cândido Costa Pires, piso 1
Braga
PT

Abstract

Background: Diabetes-related foot ulceration (DFU) is a common complication of diabetes, with a significant impact on survival, healthcare costs and health-related quality of life.

Objective: We aimed to identify and collect the available evidence assessing the ability of machine learning (ML) based models in predicting clinical outcomes in people with DFU.

Methods: We searched the MEDLINE database (PubMed), Scopus, Web of Science and IEEExplore for articles published up to July 2023. Studies were eligible if they were anterograde analytical studies that examined the prognostic abilities of ML models in predicting clinical outcomes in a population that included at least 80% of adults with DFU. The literature was screened independently by two investigators for eligibility criteria and extracted data. The risk of bias was evaluated using the Quality In Prognosis Studies (QUIPS) tool and Prediction Model Risk Of Bias Assessment Tool (PROBAST) by two investigators independently.

Results: We retrieved a total of 2412 references after removing the duplicates, of which 167 were subjected to full text screening. Two references were added from searching relevant studies' list of references. A total of 11 studies, comprising 13 articles, were included focusing on three outcomes: wound healing, lower extremity amputation and mortality. Overall, 55 predictive models were created using mostly clinical characteristics, random forest as the developing method and area under the curve (AUC) as discrimination accuracy measure. AUC varied from 0.56 to 0.94, with the majority of the models reporting an AUC ? 0.8. All studies were found to have a high risk of bias, mainly due to a lack of uniform variable definitions, outcome definitions and follow-up periods, insufficient sample sizes, and inadequate handling of missing data.

Conclusions: The ML-based models found in this study achieved good discrimination ability in people with DFU. However, models presented a high risk of bias meaning further studies with a stricter methodology are needed. Clinical Trial: PROSPERO CRD42022308248; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=308248

(JMIR Preprints 28/11/2024:69408)
DOI: https://doi.org/10.2196/preprints.69408

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Title page

Paper type

Review

Title of paper

Systematic review of diabetic foot ulcer classification models using artificial intelligence and machine learning techniques

Running title

Artificial intelligence to classify ulcers

Authors and institution(s) of origin

Manuel Alberto Silva, MD
USF Sanus Carandá, ULS Braga, Braga, Portugal
Faculty of Medicine, University of Porto, Porto, Portugal

Emma J Hamilton, MD
Department of Endocrinology and Diabetes, Fiona Stanley Hospital, Murdoch, Australia
University of Western Australia, Medical School, Fiona Stanley Hospital, Murdoch, Australia

David A Russell, MD, PhD
Leeds Institute of Clinical Trials Research, University of Leeds, Leeds, UK
Leeds Vascular Institute, Leeds Teaching Hospitals NHS Trust, Leeds, UK

Fran Game, MD, PhD
University Hospitals of Derby and Burton NHS Foundation Trust, Derby, UK

Sheila C Wang, MD, PhD
Division of Dermatology, Department of Medicine, University of Toronto, Toronto, ON, Canada

Sofia Baptista, MD, PhD
MEDCIDS - Departamento de Medicina da Comunidade Informação e Decisão em Saúde; Faculty of Medicine, University of Porto, Porto, Portugal

CINTESIS@RISE - Center for Health Technology and Services Research; Faculty of Medicine, University of Porto, Porto, Portugal
CUF Porto Hospital, Portugal

Matilde Monteiro-Soares, PhD
Portuguese Red Cross Health School Lisbon, Lisbon, Portugal
Cross I&D, Lisbon, Portugal
MEDCIDS - Departamento de Medicina da Comunidade Informação e Decisão em Saúde; Faculty of Medicine, University of Porto, Porto, Portugal
CINTESIS@RISE - Center for Health Technology and Services Research; Faculty of Medicine, University of Porto, Porto, Portugal

Corresponding author
Manuel Alberto Silva, MD
USF Sanus Carandá, ULS Braga, Braga, Portugal
Adress: Praça Cândido Costa Pires, piso 1, 4715-402 Braga, Portugal
Phone: +351253201530
E-mail: manuelalbertorsilva@gmail.com

Abstract

Background: Diabetes-related foot ulceration (DFU) is a common complication of diabetes, with a significant impact on survival, healthcare costs and health-related quality of life.

Objectives: We aimed to identify and collect the available evidence assessing the ability of machine learning (ML) based models in predicting clinical outcomes in people with DFU.

Methods: We searched the MEDLINE database (PubMed), Scopus, Web of Science and IEEExplore for articles published up to July 2023. Studies were eligible if they were anterograde analytical studies that examined the prognostic abilities of ML models in predicting clinical outcomes in a population that included at least 80% of adults with DFU. The literature was screened independently by two investigators for eligibility criteria and extracted data. The risk of bias was evaluated using the Quality In Prognosis Studies (QUIPS) tool and Prediction Model Risk Of Bias Assessment Tool (PROBAST) by two investigators independently.

Results: We retrieved a total of 2412 references after removing the duplicates, of which 167 were subjected to full text screening. Two references were added from searching relevant studies' list of references. A total of 11 studies, comprising 13 articles, were included focusing on three outcomes: wound healing, lower extremity amputation and mortality. Overall, 55 predictive models were created using mostly clinical characteristics, random forest as the developing method and area under the curve (AUC) as discrimination accuracy measure. AUC varied from 0.56 to 0.94, with the majority of the models reporting an AUC $\geq$ 0.8. All studies were found to have a high risk of bias, mainly due to a lack of uniform variable definitions, outcome definitions and follow-up periods, insufficient sample sizes, and inadequate handling of missing data.

Conclusions: The ML-based models found in this study achieved good discrimination ability in people with DFU. However, models presented a high risk of bias meaning further studies with a stricter methodology are needed.

Trial Registration: PROSPERO CRD42022308248;
https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=308248

Key-words: Artificial intelligence; diabetic foot; classification; machine learning; prognosis.

Introduction

Diabetes is a rapidly growing disease. Since 2000, the prevalence of diabetes has more than tripled, reaching, in 2021, 10.5% of the adult population in the world. [1]

The increase in diabetes prevalence is associated with the rise of its related complications. [2] Diabetes-related foot ulceration (DFU), defined as a break in the skin of the foot that involves at least the epidermis and part of the dermis [3], is the most commonly recognised complication affecting the lower extremities. The risk of a person with diabetes developing a DFU across their lifetime is around 19 to 34% [4]. Approximately 20% of people who develop a DFU will require lower-extremity amputation (LEA) [4], and 10% will die within one year of their first DFU diagnosis [5, 6]. In the United States, foot complications contribute to $273 billion in direct costs and $90 billion in indirect costs [7]. Apart from the impact of a DFU on mortality and healthcare costs, people with DFUs also have a significantly lower health-related quality of life [8].

The evaluation and prognosis of a DFU varies considerably according to person, limb and ulcer-related characteristics. For that reason, classification and scoring systems were developed to create groups of patients with similar characteristics for whom similar levels of care would apply. Furthermore, they can be used to communicate wound and person characteristics between professionals, estimate an individual's prognosis, help in clinical practice decision-making, and audit and comparison of populations.

A systematic review from the International Working Group on the Diabetic Foot (IWGDF) in 2023 found 28 different classification and scoring systems for DFUs [9]. No gold standard exists, despite the many classification and scoring systems available. Therefore, each system should be used according to the intended purpose, available resources, expertise, and clinical setting. In the IWGDF 2023 updated guidelines [10] several recommendations were made: SINBAD system (Site, Ischemia, Neuropathy, Bacterial infection, Area and Depth) was recommended for communication between healthcare professionals and clinical audit; IDSA/IWGDF (Infectious Diseases Society of America/International Working Group on the Diabetic Foot) was recommended for infected ulcers; and WIfI (Wound, Ischemia, foot Infection) was recommended for peripheral arterial disease (PAD), or for when resources and expertise are available. No available classification and scoring system for individual prognostication was recommended.

Expert opinion and conventional statistical methods, such as linear regression and other generalised linear models, have been used to develop classifications to help predict clinical outcomes in people with DFU [11, 12]. However, these methods lack detail and do not capture the complex non-linear relationships between risk factors and outcomes, compromising the classification systems' predictive

ability.

Recent technological advances have allowed the use of machine learning (ML) strategies, a branch of artificial intelligence, in healthcare. ML algorithms can use data from several sources, capture complex patterns and thus may perform better than traditional models [13, 14], especially in settings with high variability.

ML algorithms can be divided into four categories: supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning [15, 16]. In supervised learning, the system infers a function from labelled training data and a collection of training examples. The most common are classification models (predict classes) and regression models (predict numerical values) [15, 16]. In unsupervised learning, the system tries to discover the hidden structure of data or associations between variables without the need for human interference, commonly using clustering and association [15, 16]. In semi-supervised learning, there is a combination of supervised and unsupervised methods, and the system works with labelled and unlabelled data [15, 16]. In reinforcement learning, the system tries to learn through direct interaction with the environment [15, 16].

ML has been applied successfully to healthcare. A systematic review from Kavakiotis and colleagues searched the applications of ML in diabetes research and found that most of the algorithms (85%) used supervised approaches, usually when performing prediction tasks [17]. When it comes to DFU care, a systematic review by Tulloch and colleagues found multiple applications of ML, namely in classification, image analysis and segmentation [18]. However, this review focused on identifying the presence and type of DFU but not predicting clinical outcomes in people with DFUs.

Our systematic review aimed to collect all the available evidence assessing the prognostic abilities of ML-based models in predicting clinical outcomes in people with DFUs. We focused on the comparison between models, their performance and discussed their applicability in the DFU care context. We hope they can facilitate decision-making and debate the importance of integrating this type of model into daily clinical practice worldwide.

Methods

This systematic review was conducted using the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) [19] guidelines, and we used the AMSTAR (A MeaSurement Tool to Assess systematic Reviews) [20] tool to verify if the most important aspects have been included in our systematic review. We registered our review in the PROSPERO (International Prospective Register of Systematic Reviews) database in July 2022 and updated in August 2023 under

CRD42022308248.

Search strategy

We searched the MEDLINE database (PubMed), Scopus, Web of Science and IEEExplore in two phases. In the first phase, we performed a search on 26 February 2022 to identify all studies published with no beginning date until December 2021 (inclusive). In the second phase, we updated the search on 10 August 2023 to identify all studies published from January 2022 until July 2023 (inclusive). No restrictions were applied.

Search queries are available in Multimedia Appendix 1. To refine our query, we have used as "satellite" some pertinent articles included in the systematic review from Tulloch and colleagues that addressed a similar topic [18].

Reference lists of the included articles and previous systematic reviews were reviewed to find additional relevant articles. Experts in the area were contacted to identify any other articles not identified by our query, namely internal medicine physicians, vascular surgeons, endocrinologists, nurses, podiatrists, and human movement scientists.

Inclusion and exclusion criteria

Studies were selected based on the PECO-S elements (population, exposure, comparator, outcome and study type), as explained below.

Population

Articles were considered eligible if the population included at least 80% of adults with diabetes and a foot ulcer or if a subgroup analysis of such participants was provided.

Exposure/comparator

We defined the exposure of interest as being classified at higher risk (exposure) or lower risk (comparator) by any model developed using artificial intelligence techniques to predict outcomes by assessing more than one patient, foot and/or ulcer characteristic.

We also investigated the association between the models' composing variables (i.e., each variable included in the model) and the different outcomes.

Outcomes

The authors selected the outcomes for the study using the list provided in the systematic review by Dovell and colleagues as a foundation [21]. Definitions of the outcomes were made according to the document from the IWGD [3].

Our primary outcome was wound healing: reaching intact skin, meaning complete epithelialisation without any drainage of a previous foot ulcer site.

As secondary outcomes, we used the following:

a) lower-extremity amputation (LEA): resection of a segment of a lower limb through a bone or a joint;
b) hospitalisation: care in a hospital that requires admission as an inpatient and usually requires an overnight stay;
c) length of stay: period of time in which a person is committed to a hospital;
d) health-related quality of life: a person's perceived physical and mental health;
e) survival: the state or fact of continuing to live or exist;
f) ulcer-free survival period/time: period of time in which a person is alive and without a foot ulcer;
g) LEA-free period: period of time in which a person is alive and without a LEA.

Study type

We included analytic longitudinal studies, meaning clinical trials and cohort studies. If a study was presented as an abstract or poster, further searching was done to identify if it gave origin to a full article. If not, the study was excluded.

Eligibility assessment and data extraction

Articles were included if they fulfilled all the eligibility criteria mentioned above.

The search was conducted in two phases. In both phases, the studies were reviewed independently by two reviewers: Matilde Monteiro-Soares (MMS) and David Russell (DR) or Emma Hamilton (EH) in the first phase, MMS and Manuel Alberto Silva (MAS) in the second phase. Studies were selected based on their titles and abstracts in the first stage and the complete text of the articles in the second stage. Divergent opinions were resolved by consensus. We used EndNote 20® to manage references and identify duplicates. Subsequently, we used Rayyan QCRI [22] for the blind and independent selection of references to be included in our systematic review. The proportion of agreement between

the two reviewers was calculated for each stage.

Data was extracted from each included study using a spreadsheet and summarised in tables that included the following information: (i) article identification (namely authors, year of publication, country where study was conducted), (ii) methods (namely study design, inclusion of participants, sample size, follow-up, context of study), (iii) model characteristics (namely purpose, methods for development, validation conducted, variable definitions), (iv) outcome definition, and (v) results and analysis (namely participants' age, type of diabetes, diabetes duration, sex, measures and statistical methods used).

Data was extracted by one reviewer (MAS) and confirmed by a second reviewer (MMS). Divergent opinions were solved by consensus.

Risk of bias

The risk of bias was assessed using the Cochrane Risk-of-Bias (RoB 2) tool [23] for randomised trials for impact analysis. If a study had a low risk of bias in all five domains, it was classified as at low risk of bias; if some concerns existed in at least one domain without any domain with a high risk of bias, it was classified as with some concerns; and if at least one domain had a high risk of bias or some concerns existed for multiple domains, it was classified as at high risk of bias.

For observational longitudinal studies of clinical prognosis, both the Quality In Prognosis Studies (QUIPS) tool [24] and the Prediction model Risk Of Bias ASsessment Tool (PROBAST) [25] were used. In the case of QUIPS, we evaluated the studies according to five of the six proposed domains. We considered that the study confounding domain was not pertinent as the present paper aims to study the association between variables and outcomes regardless of a causal relationship. Thus, this domain was classified as low risk in all studies. Overall, if a study had a low risk of bias in the six domains, it was classified as being at very low risk of bias, in four to five domains as being at low risk of bias, and three or fewer domains as being at high risk of bias. In the case of PROBAST, if a prediction model had a low risk on all domains relating to bias and applicability, it was classified as at low risk of bias or low concern regarding applicability; if a model had a high risk for at least one domain, it should be classified as having high risk of bias or high concern regarding applicability; if a model had unclear risk in one or more domains and had low risk in the remaining domains, it may be classified as having unclear risk of bias or unclear concern regarding applicability.

Two reviewers (MMS and MAS) assessed the risk of bias. Divergent opinions were resolved by consensus. The proportion of agreement between the two reviewers was calculated.

Results

Search results

In the first phase, we retrieved a total of 1950 references after removing the duplicates. A total of 90 references were selected in the first stage (title and abstract screening), with a proportion of agreement of 95% among the assessors (MMS, DR and EH). After the second stage (full-text screening), with an agreement of 98%, four references were included in our systematic review.

In the second phase (update of search), we retrieved 462 additional references. A total of 81 references were selected in the first stage, with a proportion of agreement of 90% between the two assessors (MMS and MAS). After the second stage, we included seven additional references in our systematic review, with an agreement of 98%. From searching the references of previous reviews (systematic or not), of included studies, and from contacting experts, we retrieved an additional two references. Thus, we included 13 papers reporting on 11 studies (see Figure 1).

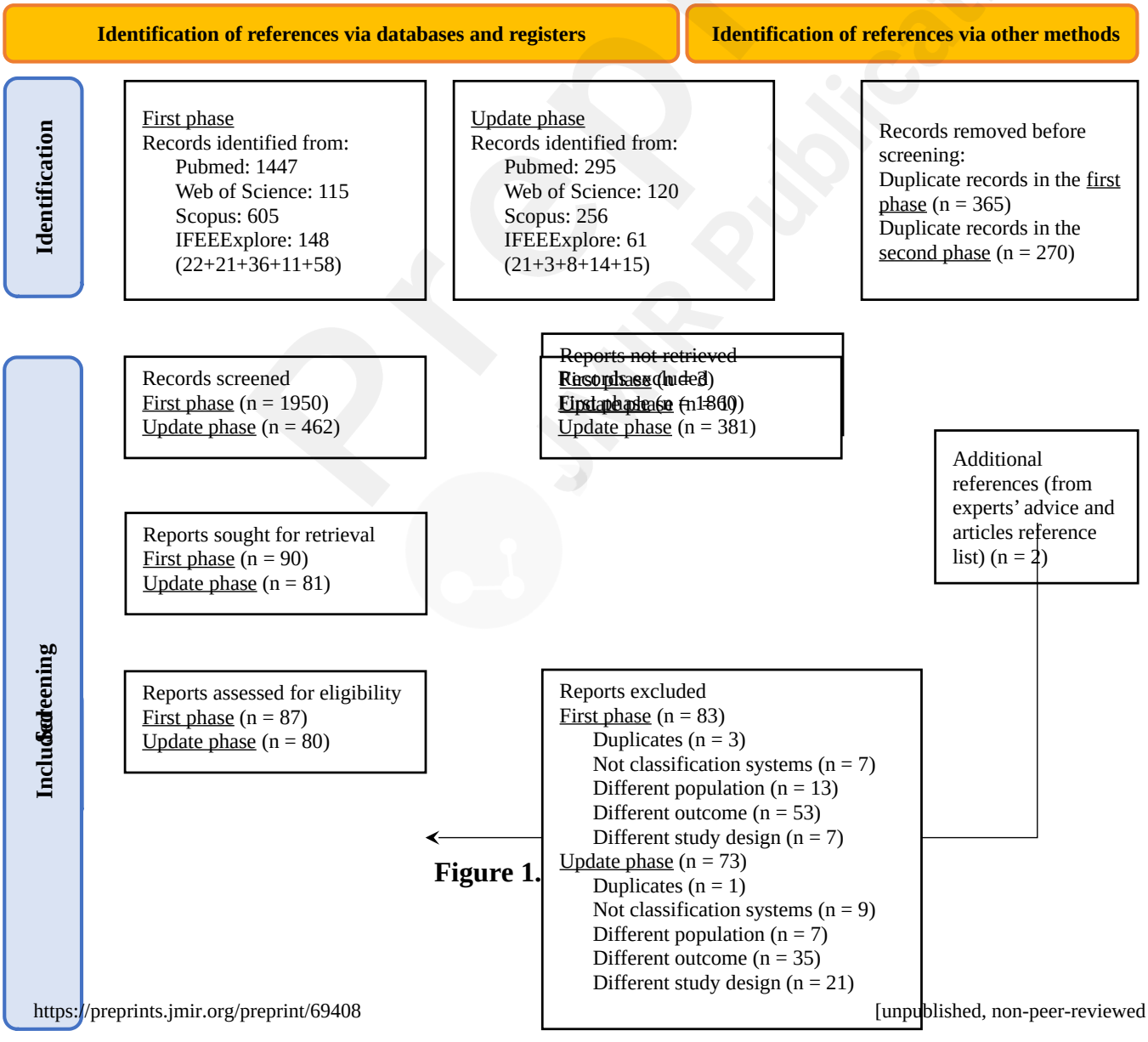


Figure 1.

Studies' design, setting and population

We have included studies published between 2016 and 2023, conducted in five countries: the United States (n=5, 45%) [26-30], China (n=3, 27%) [31-34], Germany (n=1, 9%) [35, 36], Poland (n=1, 9%) [37], and India (n=1, 9%) [38]. Eight studies (73%) were retrospective cohorts [26, 28-34, 38], and three were prospective cohorts [27, 35-37]. Six studies (55%) were single-centre [29, 31, 34-38], with sample sizes ranging from 46 to 618 participants (median of 201), while five studies were multicentre [26-28, 30, 32, 33], and had a sample size varying from 204 to 88,898 participants (median of 53,354).

We have separated the results by the clinical outcome (wound healing, LEA and mortality) and organised the studies included in each table (Multimedia Appendices 2, 4 and 6) by a higher stage of development (meaning external validation, internal validation or derivation only), study design with less risk of bias (meaning prospective cohort, or retrospective cohort study), multicentre versus single centre, and larger sample size.

Prediction models' characteristics by clinical outcome

Wound healing

A total of five papers (38%) [27, 29, 30, 32, 37] used wound healing as an outcome with some variations (see Multimedia Appendix 2): two studies [27, 29] evaluated wound healing; one study assessed delayed wound healing [30]; one study evaluated hard-to-heal wound [32]; and one study [37] evaluated wound healing failure. Follow-up periods for the mentioned outcomes ranged from four to 16 weeks, while one study [29] did not explicitly define any period of time for measuring the outcome.

Four studies [27, 29, 32, 37] included only DFUs, with a total of 846 participants. Jung et al. [30] included several types of wounds, with 6,055 (4% of the overall sample) being neuropathic DFUs and provided a subgroup analysis. The incidence of wound healing varied from 35.1 to 78.8%. Regarding hard-to-heal wounds and wounds with delayed or failed healing, the incidence ranged from 11.6 to 66.0%.

A median of 35 clinical variables per study were assessed for model construction. Final models included between four and 865, with a median of 10 variables per study, distributed across four categories: demographic characteristics, medical history, laboratory data and foot-related characteristics (see Multimedia Appendix 3). Kim et al. [29] also included image-based characteristics retrieved from photographs through the user's subjective observation and deep learning techniques. The most commonly included were wound area (n=4, 80%), sex (n=3, 60%) and C-reactive protein (CRP) (n=3, 60%).

Regarding ML methods, four studies (80%) [29, 30, 32, 37] employed multiple techniques simultaneously. The most applied ML method was random forest (RF), which was used in four studies (80%), followed by support vector machine (SVM) and least absolute shrinkage and selection operator (LASSO) regression, which was used in two studies each (40%).

Overall, across the five papers, 20 prediction models were created. Missing data was excluded in two studies [30, 32], and was not reported in one study [37]. One study [27] reported the existence of missing values but not the approach to handle them, and one study [29] used imputation with a k-nearest neighbors (k-NN) algorithm. As for model validation, every study used internal validation processes. Calibration was evaluated in two studies [27, 30] using the Hosmer-Lemeshow test and Brier score, and discrimination accuracy was assessed by several measures including the area under the curve (AUC). Reported AUCs in model testing ranged from 0.636 to 0.864. The model which showed apparent better discrimination was developed in the study by Wang and colleagues, [32] using 10 clinical variables and the naïve Bayesian classifier. However, 95% confidence intervals (CI) were never reported.

Lower extremity amputation

Eight articles (54%) [26, 28, 31, 33-36, 38] had LEA as outcome (see Multimedia Appendix 4). One article focused on major LEA [26], another article on minor LEA [33], and four articles assessed simultaneously two different types of LEA (minor and major, and major and any) [28, 31, 35, 36]. The remaining papers focused on any form of amputation. The follow-up period for determining LEA occurrence ranged from six to 12 months; four studies [26, 31, 33, 34] did not mention a predefined follow-up period for determining the outcome.

A total of 417,315 people with diabetes were included. The incidence of major LEA ranged from 5.9 to 12.2%, whereas of minor LEA ranged from 11.5 to 20.7%. Concerning any form of LEA, the incidence varied from 1.6 to 31.6%.

A median of 21 clinical variables per study were assessed for model construction. Final models

included between seven and 37, with a median of 10 variables per study. Du et al. [34] only reported the most relevant variables to model construction, so only those were accounted for. The most frequent variables were age (n=5, 71%), and sex, diabetes duration, smoking history, haemoglobin A1c, creatinine, albumin and random blood glucose (all n=3, 43%) (see Multimedia Appendix 5).

Regarding ML techniques, four studies (57%) employed multiple techniques simultaneously [26, 28, 33, 34]. The most used ML method was RF, which was used in four studies (57%) [26, 28, 33, 34].

Overall, across the seven papers, 27 prediction models were created. Missing data was inappropriately handled in two studies [26, 33], and was not reported in the remaining studies. As for model validation, every study, apart from one (that led to two articles) [35, 36], used internal validation processes. Calibration was evaluated in two studies [28, 31] using the McFadden R^2, isotonic regression and Brier score, and discrimination accuracy was assessed through AUC in five papers [26, 31, 33-36]. Reported AUCs ranged from 0.60 to 0.90. The model which showed apparent better discrimination ability was developed in the study by Xie and colleagues [31], using 37 clinical variables and a Light Gradient Boosting Machine (Light GBM). One study reported an accuracy of 94% [38], and one study [28] only reported an out-of-bag error rate, which varied from 31 to 63%. Kaskebar et al. and Husers et al. reported 95% CI, allowing comparisons between models.

Mortality

Mortality was defined as an outcome in two articles (15%) [28, 34], and, in one of them [28], it was measured after six months (see Multimedia Appendix 6). A total of 88,944 persons were included. The mortality rate varied from 4.5 to 17.4%.

The final models included a median of 10 variables per study, with Du et al. [34] only describing the most relevant variables for model construction. The only repeated variable was age (n=2, 100%) (Multimedia Appendix 7).

Both studies used multiple ML techniques including logistic regression (LR) and RF in their analysis.

Overall, eight prediction models were created. Neither of the studies reported missing data. Both studies used processes of internal validation. Calibration was evaluated in one study using the McFadden R^2, and discrimination accuracy was assessed, by one paper [34], using the AUC, with reported values ranging from 0.56 to 0.94. The other paper [28] only reported out-of-bag error rate varying from 30 to 68%.

Risk of bias

We evaluated the risk of bias according to two tools: QUIPS (see Multimedia Appendix 8), with a proportion of agreement of 71%, and PROBAST (see Multimedia Appendix 9), with a proportion of agreement of 88% for risk of bias and 61% for applicability.

Risk of bias according to QUIPS

All studies presented a high risk of bias and had two or three (out of the six domains) classified as being at low risk (see Multimedia Appendix 8).

In the study participation domain, all 11 studies were classified as having a moderate or high risk of bias, mainly because most studies did not clearly explain how participants' sampling was conducted and failed to describe fully eligibility criteria.

In the study attrition domain, nine studies (82%) were classified as having a high risk of bias. Of these, seven did not mention the proportion of patients that concluded the study [26, 28, 31-36, 38], and two [29, 30], although reported the response rate, did not characterise the excluded patients or the reasons for their exclusion.

As for the prognostic factor measurement domain, all studies were classified as having moderate or high risk of bias. Nine studies (82%) [26, 27, 29-34, 37, 38] failed to clearly define or explain how to collect all the analysed variables. Several variables, such as wound area, wound depth, erythema and adequate arterial flow, are subjective and can lead to different results if measured differently. Two studies [27, 30] may have introduced bias due to using several patient centres with no standardised protocols provided. Apart from one study [29], in which imputation with a k-NN algorithm was used, missing data was not reported, not addressed or incorrectly handled.

Concerning the outcome measurement domain, four studies (36%) were classified as having a moderate or high risk of bias. Three of these studies [27, 29, 30] failed to define the outcome clearly, and none specified a protocol for outcome assessment.

Statistical analysis and reporting were considered in all studies to be at low risk of bias.

Risk of bias according to PROBAST

All studies presented a high risk of bias, and 10 presented high or unclear concerns for applicability (see Multimedia Appendix 9).

Five studies were classified as having high or unclear risk of bias in the participants' domain due to inadequate or absent eligibility criteria. In the evaluation of applicability, seven studies had a high or unclear concern for applicability. Similarly, this is mainly due to the inclusion of participants that

may differ from the studies' target population.

In the predictors domain, nine studies were classified as having unclear risk of bias and unclear risk of applicability. Eight studies [26, 28-34, 38] had a retrospective design and failed to mention if the predictors' assessment was made without the knowledge of outcome data. Two studies [27, 30] included multiple patient centres, and a homogenous assessment of predictors was not clearly guaranteed.

As for the outcomes, nine papers were classified as having a high or unclear risk of bias. In four of these [27, 29, 30, 37], authors failed to report a clear definition for the outcome, compromising its measurement. Another factor that potentially introduced bias was the lack of follow-up reporting (n=5) [26, 31, 33, 34, 38]. In the evaluation of applicability, four studies had a high concern for applicability due to the lack of clear definitions for the outcomes.

In the analysis domain, all studies were classified as having a high risk of bias. Nine studies had insufficient sample sizes, with only two studies [26, 28] reaching over 200 events per variable (EPV) tested – the cut-off considered reasonable to minimise overfitting when using ML techniques. Three studies [26, 30, 32, 33] converted continuous variables into categorical ones, using arbitrary rules for categorisation. Regarding missing data, only one study [29] addressed missing data correctly, using imputation with a k-NN algorithm. Two studies [32, 33, 37] failed to avoid selecting variables solely based on univariate analysis, whereas two [30, 38] did not mention how variables were selected. As for model evaluation, seven studies [26, 29, 32-38] did not report any calibration. The remaining used different methods, including Hosmer-Lemeshow statistics, the Brier score, the McFadden R^2 and the isotonic regression. Discrimination accuracy measures were employed by all studies except one [28], with AUC being the most often used. However, uncertainty measures were seldom provided. In terms of validation, apart from one (that only derived models) [35, 36], every study relied solely on internal validation methods.

Discussion

Principal findings

DFUs have become one of the most important causes of mortality and morbidity in people with diabetes. Guideline-directed foot exams and treatments and aggressive medical management of diabetes and cardiovascular disease are paramount in improving the prognosis of people with a DFU. However, effective interventions do not work in the same way in all people, and the same management cannot be provided to everyone with diabetes. For that reason, predictive models have been used to stratify people with DFU by their probability of healing, requiring an LEA and dying,

so that interventions and resources can be appropriately allocated.

Our systematic review aimed to ascertain if models using an ML approach and clinical data that are easily accessible in clinical practice could predict clinical outcomes. We have included 11 studies corresponding to 13 articles, mainly from the USA and China (73%), retrospective (73%), single centre (55%), and with LEA as the outcome (54%).

Other reviews have previously investigated how ML could improve DFU care. For example, a systematic review by Tulloch and colleagues [18] focused on identifying and classifying DFU at a specific moment. In this review, the predictive abilities of ML algorithms were not considered, unlike ours. More recently, a systematic review by Huang and colleagues [39] searched the literature for models that predicted amputation. This study focused on prognosis but considered models that used several methodologies (ML and not ML) and that predicted only one outcome (LEA). This review included people with “diabetic foot”, a broader concept than DFU, characterised by ulcers or destruction of the tissues due to infection and/or peripheral artery disease.

Although we have included 11 studies (published as 13 articles), we found a total of 55 prediction models focusing on three outcomes: healing (five articles presented 20 models), LEA (seven articles presented 27 models), and mortality (two articles presented eight models) . Our search did not retrieve some of the outcomes defined in our protocol, such as hospitalisation, length of stay, health-related quality of life, ulcer-free survival period/time and LEA-free period, nor reliability studies, external validation studies, nor studies assessing the impact of developing and implementing a predictive model in clinical practice.

For model construction, studies used clinical variables distributed into four categories: demographic characteristics, medical history, laboratory data and foot-related characteristics. The number of variables in the final models varied from four to 865.

When it comes to wound healing, variables from the foot-related characteristics group were the most frequently selected, with wound area being used in models from almost all studies. In fact, in previous studies, ischemic ulcers, more extensive and deeper ulcers, plantar ulcers, and ulcers with infection have been associated with poor healing [40, 41]. Besides clinical variables, Kim et al. [29] also included image-based characteristics from photographs through subjective and deep learning analysis. The subjective observation allowed models to be adequately trained with good performance, turning the utilisation of smartphone and tablet photographs for prognosis assessment in clinical practice into a possibility.

Regarding LEA, the most frequently selected variables were age, sex, diabetes duration, smoking history, haemoglobin A1c, creatinine, albumin and random blood glucose. In the case of sex, males

have been reported to have higher amputation rates, likely due to behavioural differences [42, 43]. As for smoking, studies have suggested its association with LEA by increasing the risk of atherosclerosis and, consequently, of PAD [42, 43]. Decreased albumin levels have also been connected to higher LEA rates [43]. As for the remaining variables, a recent systematic review with meta-analysis found no correlation between these and the risk of LEA [43], questioning the methods chosen by these papers to select predictors for model construction.

Lastly, in the case of mortality, age was the only variable repeated in the two selected studies. Despite this being an important outcome, not many studies have sought to create models that could predict mortality.

The most used ML method was RF. It was first described by Breiman et al. [44], and is a supervised method that uses “parallel ensembling”, fitting several decision tree classifiers in parallel, where each tree is trained on a random subset of the training data with replacement (bagging). Majority voting or averages are used to obtain the final result. This method is suitable for both classification and regression problems and has a reduced risk of overfitting, when compared to decision trees. However, it is a time-consuming process, that requires more resources, and with a more complex interpretation [16].

The most reported discrimination accuracy measure was AUC. It refers to the area under the receiver operating characteristic (ROC) curve and is an effective way of summarising the overall diagnostic accuracy of a model, taking values from zero to one. The ROC curve is depicted by using each possible value of a continuous variable as a point with a certain sensitivity and (1-) specificity in discrimating those with and without the clinical condition of interest [45].

There is some variation in the qualitative descriptors of model performance for AUC thresholds. According to Mandrekar et al. [45], a AUC value of 0.7 to 0.8 is considered acceptable, 0.8 to 0.9 is considered excellent, and more than 0.9 is considered outstanding. However, the AUC is a combined measure of the overall sensitivity and specificity of a model. This implies that two models can have identical AUC values, and one perform better in higher sensitivities and the other perform better in higher specificities [46]. Futhermore, AUC values may overestimate model performance, when the same database is used for both testing and training as happened in most studies. Therefore, interpretation of AUC values must be done cautiously.

Overall, most models were able to achieve good discrimination ability, with 51% of the reported AUC values being equal or superior to 0.8. However, all studies were considered to have a high risk of bias, according to QUIPS and PROBAST. Firstly, most studies failed to clearly describe inclusion and/or exclusion criteria, raising doubt about the potential applicability of final models. Patients lost

to follow-up, when reported, were also excluded. Secondly, most studies did not clearly define the variables considered or describe the methodology adopted to measure them. Some variables, such as those in the foot-related characteristics group, were subjective, and different assessors may measure them in various ways. As a result, without clear definitions and established protocols for variables' measurement, predictive models' validity and application to clinical practice may be compromised. This lack of standardisation also applies to outcomes, where clear definitions are essential to guarantee consistency. In the case of the studies considering wound healing as an outcome, different variations were used (delayed wound healing, hard-to-heal wound, and wound healing failure). However, most did not provide a definition or a methodology for outcome measurement. Also, follow-up periods varied widely among studies, which makes comparing predictive models even more challenging. These variations and lack of standardisation were expected; thus, no meta-analysis was considered from inception. Thirdly, there were some problems in the analysis domain. Studies had insufficient sample sizes, with only two reaching over 200 EPV [26, 28]. ‘EPV’ refers to the number of events (i.e., number of patients in which the outcome of interest has occurred) relative to the number of regression coefficients used (i.e., number of variables considered for model devolpment) [25]. In ML-based models, higher EPV (often more than 200) are needed to minimize overfitting [47]. Also, most studies used the same sample set to develop (train) and validate (test) the models, which can increase the overestimation of the accuracy due to overfitting. Missing data was not reported, not addressed or incorrectly handled, except in one study that used imputation methods. Although reported models showed promise in predicting clinical outcomes, most are not available for immediate applicability to clinical practice. Only three studies presented online interactive models: two [26, 33] for LEA prediction and one [32] for hard-to-heal DFU prediction. In addition, several studies developed and internally validated several models using different methods in the same sample without providing all required accuracy measures (focusing on AUC) and seldomly reporting 95% confidence intervals – making it impossible to compare and select specific models to be externally validated before they can be used in clinical practice.

Limitations

Our systematic review presents some limitations. Firstly, we conducted our search in two phases, involving different reviewers in the selection process, which may have introduced some variability. However, the agreement proportion was high and comparable when comparing both phases. Secondly, a systematic review is only as good as the included studies. Due to the high risk of bias and the high heterogeneity of included studies (mainly due to a lack of uniform variable definitions, outcome definitions and follow-up periods), we have decided that a meta-analysis should not be

performed. However, we would like to highlight that this reality is not as different from the results achieved in our systematic review of “classic” classification systems. Lastly, the complexity of the ML methodology prevented further explanation of the studied models, which may lead to some distrust from healthcare providers when considering the application of such models in clinical practice.

Future reasearch

For future research, studies with larger samples with a low risk of bias, preferably in an external validation setting, are needed. When developing DFU prediction models, we recommend that variables be selected according to the available evidence provided by previous studies. Additionally, other outcomes, such as quality of life, which are clinically relevant, should be further utilised.

Conclusion

This systematic review found several ML-based models that could predict clinical outcomes with good discrimination ability in people with DFU, showing promising results. However, studies presented a high risk of bias with several applicability issues that compromise the ready applicability of such models in clinical practice. Further studies with stricter methodology are needed so that patients who have diabetes and a foot ulcer can benefit from the recent advancements of artificial intelligence applied to healthcare.

Acknowledgments

Publication of the study was supported by Faculty of Medicine, University of Porto.

Authors' contributions

MMS, DR and EH formulated the research question, the inclusion and exclusion criteria, designed search terms, and screened the records in the first phase.

MMS and MAS screened the records in the second phase, performed data extraction, assessed risk of bias, drafted and revised the initial version of the manuscript, and are responsible for the decision to submit the manuscript for publication.

All the authors interpreted the data, revised the manuscript, and approved the final version of the manuscript, and had full access to all the data in the study.

Conflicts of interest

David Russell is supported in part by the National Institute for Health and Care Research (NIHR) Leeds Biomedical Research Centre (BRC) (NIHR203331). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care. None of the remaining authors have conflicts of interest to declare.

Abbreviations:

AMSTAR: A MeaSurement Tool to Assess systematic Reviews

AUC: Area Under the Curve

CI: Confidence Intervals

CRP: C-Reactive Protein

DFU: Diabetes-related Foot Ulceration

DR: David Russell

EH: Emma Hamilton

EPV: Events Per Variable

IDSA: Infectious Diseases Society of America

IWGDF: International Working Group on the Diabetic Foot

k-NN: k-nearest neighbors

LASSO: Least Absolute Shrinkage and Selection Operator

LEA**:** Lower-Extremity Amputation

Light GBM: Light Gradient Boosting Machine

LR: Logistic Regression

MAS: Manuel Alberto Silva

ML: Machine Learning

MMS: Matilde Monteiro-Soares

PAD: Peripheral Arterial Disease

PECO-S: Population, Exposure, Comparator, Outcome and Study type

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analysis

PROBAST: Prediction model Risk Of Bias ASsessment Tool

PROSPERO: International Prospective Register of Systematic Reviews

QUIPS: Quality In Prognosis Studies

RF: Random Forest

ROC: Receiver Operating Characteristic

SINBAD: Site, Ischemia, Neuropathy, Bacterial infection, Area and Depth

SVM: Support Vector Machine

WIfI: Wound, Ischemia, foot Infection

References

1. IDF Diabetes Atlas. 2021. URL: https://diabetesatlas.org/atlas/tenth-edition/ [accessed june 29, 2024]
2. Harding JL, Pavkov ME, Magliano DJ, Shaw JE, Gregg EW. Global trends in diabetes complications: a review of current evidence. Diabetologia. 2019 Jan;62(1):3-16. PMID: 30171279. doi: 10.1007/s00125-018-4711-2.
3. van Netten JJ, Bus SA, Apelqvist J, Lipsky BA, Hinchliffe RJ, Game F, et al. Definitions and criteria for diabetic foot disease. Diabetes Metab Res Rev. 2020 Mar;36 Suppl 1:e3268. PMID: 31943705. doi: 10.1002/dmrr.3268.
4. Armstrong DG, Boulton AJM, Bus SA. Diabetic Foot Ulcers and Their Recurrence. N Engl J Med. 2017 Jun 15;376(24):2367-75. PMID: 28614678. doi: 10.1056/NEJMra1615439.
5. Hoffstad O, Mitra N, Walsh J, Margolis DJ. Diabetes, lower-extremity amputation, and death. Diabetes Care. 2015 Oct;38(10):1852-7. PMID: 26203063. doi: 10.2337/dc15-0536.
6. Meloni M, Izzo V, Giurato L, Lazaro-Martinez JL, Uccioli L. Prevalence, Clinical Aspects and Outcomes in a Large Cohort of Persons with Diabetic Foot Disease: Comparison between Neuropathic and Ischemic Ulcers. J Clin Med. 2020 Jun 8;9(6). PMID: 32521700. doi: 10.3390/jcm9061780.
7. McDermott K, Fang M, Boulton AJM, Selvin E, Hicks CW. Etiology, Epidemiology, and Disparities in the Burden of Diabetic Foot Ulcers. Diabetes Care. 2023 Jan 1;46(1):209-21. PMID: 36548709. doi: 10.2337/dci22-0043.
8. Khunkaew S, Fernandez R, Sim J. Health-related quality of life among adults living with diabetic foot ulcers: a meta-analysis. Qual Life Res. 2019 Jun;28(6):1413-27. PMID: 30565072. doi: 10.1007/s11136-018-2082-2.
9. Monteiro-Soares M, Hamilton EJ, Russell DA, Srisawasdi G, Boyko EJ, Mills JL, et al. Classification of foot ulcers in people with diabetes: A systematic review. Diabetes Metab Res Rev. 2024 Mar;40(3):e3645. PMID: 37132179. doi: 10.1002/dmrr.3645.
10. Monteiro-Soares M, Hamilton EJ, Russell DA, Srisawasdi G, Boyko EJ, Mills JL, et al. Guidelines on the classification of foot ulcers in people with diabetes (IWGDF 2023 update). Diabetes Metab Res Rev. 2023 May 14:e3648. PMID: 37179483. doi: 10.1002/dmrr.3648.
11. Won SH, Chung CY, Park MS, Lee T, Sung KH, Lee SY, et al. Risk factors associated with amputation-free survival in patient with diabetic foot ulcers. Yonsei Med J. 2014 Sep;55(5):1373-8. PMID: 25048499. doi: 10.3349/ymj.2014.55.5.1373.
12. Lu Q, Wang J, Wei X, Wang G, Xu Y. Risk Factors for Major Amputation in Diabetic Foot Ulcer Patients. Diabetes Metab Syndr Obes. 2021;14:2019-27. PMID: 33976562. doi: 10.2147/DMSO.S307815.
13. Desai RJ, Wang SV, Vaduganathan M, Evers T, Schneeweiss S. Comparison of Machine Learning Methods With Traditional Models for Use of Administrative Claims With Electronic Medical Records to Predict Heart Failure Outcomes. JAMA Network Open. 2020;3(1):e1918962-e. doi: 10.1001/jamanetworkopen.2019.18962.
14. Premsagar P, Aldous C, Esterhuizen TM, Gomes BJ, Gaskell JW, Tabb DL. Comparing conventional statistical models and machine learning in a small cohort of South African cardiac patients. Informatics in Medicine Unlocked. 2022 2022/01/01/;34:101103. doi:

https://doi.org/10.1016/j.imu.2022.101103.
15. Mohammed M, Khan M, Bashier E. Machine Learning: Algorithms and Applications2016. ISBN: 9781498705387.
16. Sarker IH. Machine Learning: Algorithms, Real-World Applications and Research Directions. SN Comput Sci. 2021;2(3):160. PMID: 33778771. doi: 10.1007/s42979-021-00592-x.
17. Kavakiotis I, Tsave O, Salifoglou A, Maglaveras N, Vlahavas I, Chouvarda I. Machine Learning and Data Mining Methods in Diabetes Research. Comput Struct Biotechnol J. 2017;15:104-16. PMID: 28138367. doi: 10.1016/j.csbj.2016.12.005.
18. Tulloch J, Zamani R, Akrami M. Machine Learning in the Prevention, Diagnosis and Management of Diabetic Foot Ulcers: A Systematic Review. IEEE Access. 2020;8:198977-9000. doi: 10.1109/ACCESS.2020.3035327.
19. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ. 2015 Jan 2;350:g7647. PMID: 25555855. doi: 10.1136/bmj.g7647.
20. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol. 2007 Feb 15;7:10. PMID: 17302989. doi: 10.1186/1471-2288-7-10.
21. Dovell G, Staniszewska A, Ramirez J, Murray I, Ambler GK, Twine CP, et al. A systematic review of outcome reporting for interventions to treat people with diabetic foot ulceration. Diabet Med. 2021 Oct;38(10):e14664. PMID: 34324741. doi: 10.1111/dme.14664.
22. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev. 2016 Dec 5;5(1):210. PMID: 27919275. doi: 10.1186/s13643-016-0384-4.
23. Sterne JAC, Savovic J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ. 2019 Aug 28;366:l4898. PMID: 31462531. doi: 10.1136/bmj.l4898.
24. Hayden JA, van der Windt DA, Cartwright JL, Cote P, Bombardier C. Assessing bias in studies of prognostic factors. Ann Intern Med. 2013 Feb 19;158(4):280-6. PMID: 23420236. doi: 10.7326/0003-4819-158-4-201302190-00009.
25. Moons KGM, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. Ann Intern Med. 2019 Jan 1;170(1):W1-W33. PMID: 30596876. doi: 10.7326/M18-1377.
26. Stefanopoulos S, Qiu Q, Ren G, Ahmed A, Osman M, Brunicardi FC, et al. A Machine Learning Model for Prediction of Amputation in Diabetics. J Diabetes Sci Technol. 2022 Dec 8:19322968221142899. PMID: 36476059. doi: 10.1177/19322968221142899.
27. Margolis DJ, Mitra N, Malay DS, Mirza ZK, Lantis JC, Lev-Tov HA, et al. Further evidence that wound size and duration are strong prognostic markers of diabetic foot ulcer healing. Wound Repair Regen. 2022 Jul;30(4):487-90. PMID: 35470507. doi: 10.1111/wrr.13019.
28. Austin AM, Ramkumar N, Gladders B, Barnes JA, Eid MA, Moore KO, et al. Using a cohort study of diabetes and peripheral artery disease to compare logistic regression and machine learning via random forest modeling. BMC Med Res Methodol. 2022 Nov 23;22(1):300. PMID: 36418976. doi: 10.1186/s12874-022-01774-8.
29. Kim RB, Gryak J, Mishra A, Cui C, Soroushmehr SMR, Najarian K, et al. Utilization of smartphone and tablet camera photographs to predict healing of diabetes-related foot ulcers. Comput Biol Med. 2020 Nov;126:104042. PMID: 33059239. doi: 10.1016/j.compbiomed.2020.104042.
30. Jung K, Covington S, Sen CK, Januszyk M, Kirsner RS, Gurtner GC, et al. Rapid identification of slow healing wounds. Wound Repair Regen. 2016 Jan-Feb;24(1):181-8. PMID: 26606167. doi: 10.1111/wrr.12384.
31. Xie P, Li Y, Deng B, Du C, Rui S, Deng W, et al. An explainable machine learning model for

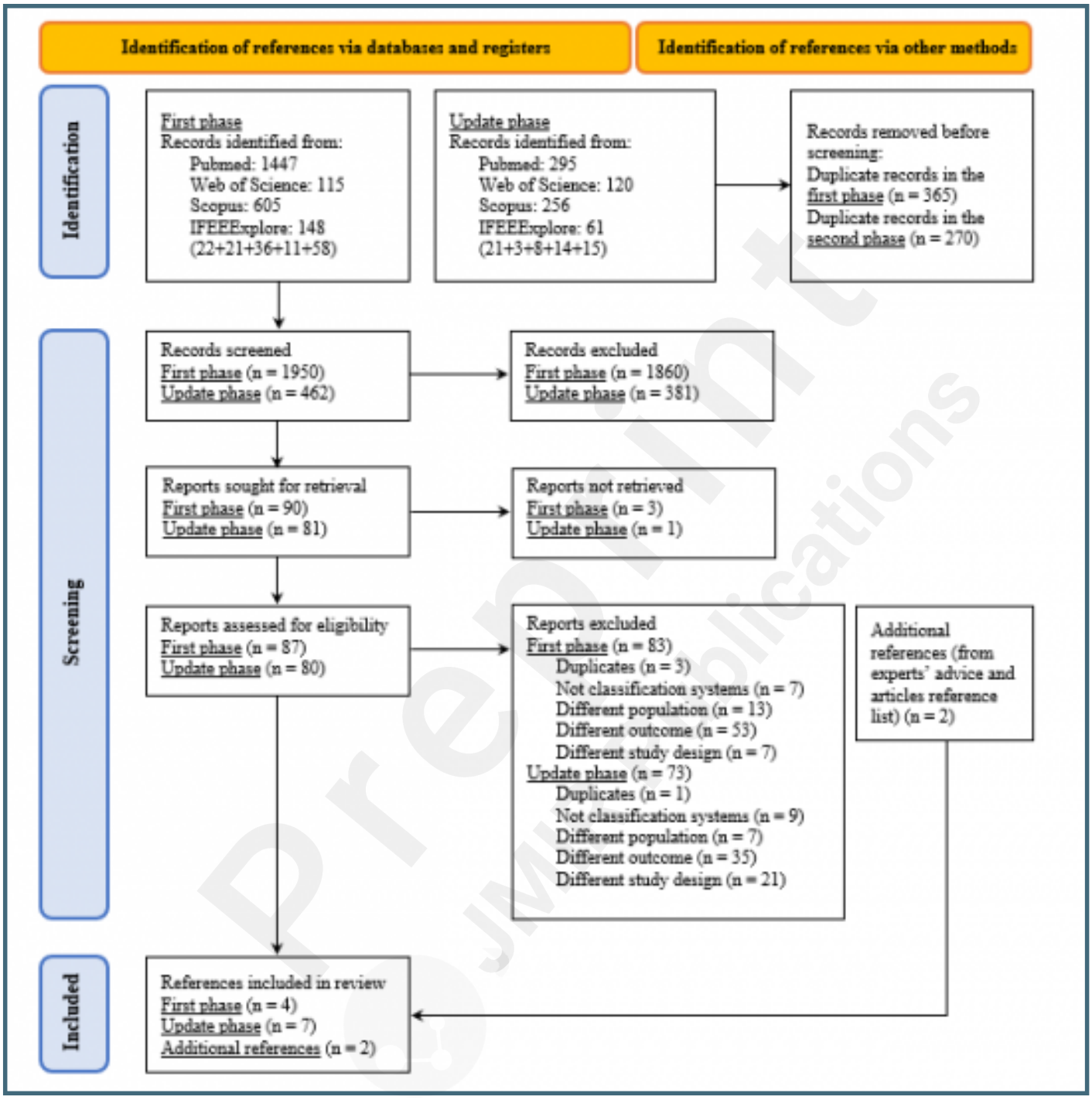
predicting in-hospital amputation rate of patients with diabetic foot ulcer. Int Wound J. 2022 May;19(4):910-8. PMID: 34520110. doi: 10.1111/iwj.13691.
32. Wang S, Xia C, Zheng Q, Wang A, Tan Q. Machine Learning Models for Predicting the Risk of Hard-to-Heal Diabetic Foot Ulcers in a Chinese Population. Diabetes Metab Syndr Obes. 2022;15:3347-59. PMID: 36341229. doi: 10.2147/DMSO.S383960.
33. Wang S, Wang J, Zhu MX, Tan Q. Machine learning for the prediction of minor amputation in University of Texas grade 3 diabetic foot ulcers. PLoS One. 2022;17(12):e0278445. PMID: 36472981. doi: 10.1371/journal.pone.0278445.
34. Du C, Li Y, Xie P, Zhang X, Deng B, Wang G, et al. The amputation and mortality of inpatients with diabetic foot ulceration in the COVID-19 pandemic and postpandemic era: A machine learning study. Int Wound J. 2022 Oct;19(6):1289-97. PMID: 34818691. doi: 10.1111/iwj.13723.
35. Husers J, Hafer G, Heggemann J, Wiemeyer S, John SM, Hubner U. Predicting the amputation risk for patients with diabetic foot ulceration - a Bayesian decision support tool. BMC Med Inform Decis Mak. 2020 Aug 24;20(1):200. PMID: 32838777. doi: 10.1186/s12911-020-01195-x.
36. Husers J, Hafer G, Heggemann J, Wiemeyer S, John SM, Hubner U. Development and Evaluation of a Bayesian Risk Stratification Method for Major Amputations in Patients with Diabetic Foot Ulcers. Stud Health Technol Inform. 2022 Jan 14;289:212-5. PMID: 35062130. doi: 10.3233/SHTI210897.
37. Poradzka AA, Czupryniak L. The use of the artificial neural network for three-month prognosis in diabetic foot syndrome. J Diabetes Complications. 2023 Feb;37(2):108392. PMID: 36623424. doi: 10.1016/j.jdiacomp.2022.108392.
38. Kasbekar PU, Goel P, Jadhav SP. A Decision Tree Analysis of Diabetic Foot Amputation Risk in Indian Patients. Front Endocrinol (Lausanne). 2017;8:25. PMID: 28261156. doi: 10.3389/fendo.2017.00025.
39. Huang J, Yang J, Qi H, Xu M, Xu X, Zhu Y. Prediction models for amputation after diabetic foot: systematic review and critical appraisal. Diabetology & Metabolic Syndrome. 2024 2024/06/10;16(1):126. doi: 10.1186/s13098-024-01360-6.
40. Zhang Y, Cramb S, McPhail SM, Pacella R, van Netten JJ, Cheng Q, et al. Factors Associated With Healing of Diabetes-Related Foot Ulcers: Observations From a Large Prospective Real-World Cohort. Diabetes Care. 2021;44(7):e143-e5. doi: 10.2337/dc20-3120.
41. Yotsu RR, Pham NM, Oe M, Nagase T, Sanada H, Hara H, et al. Comparison of characteristics and healing course of diabetic foot ulcers by etiological classification: Neuropathic, ischemic, and neuro-ischemic type. Journal of Diabetes and its Complications. 2014 2014/07/01/;28(4):528-35. doi: https://doi.org/10.1016/j.jdiacomp.2014.03.013.
42. Shin JY, Roh S-G, Lee N-H, Yang K-M. Influence of Epidemiologic and Patient Behavior–Related Predictors on Amputation Rates in Diabetic Patients: Systematic Review and Meta-Analysis. The International Journal of Lower Extremity Wounds. 2017 2017/03/01;16(1):14-22. doi: 10.1177/1534734617699318.
43. Luo Y, Liu C, Li C, Jin M, Pi L, Jin Z. The incidence of lower extremity amputation and its associated risk factors in patients with diabetic foot ulcers: A meta-analysis. Int Wound J. 2024 Jul;21(7):e14931. PMID: 38972836. doi: 10.1111/iwj.14931.
44. Breiman L. Random Forests. Machine Learning. 2001 2001/10/01;45(1):5-32. doi: 10.1023/A:1010933404324.
45. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. J Thorac Oncol. 2010 Sep;5(9):1315-6. PMID: 20736804. doi: 10.1097/JTO.0b013e3181ec173d.
46. Halligan S, Altman DG, Mallett S. Disadvantages of using the area under the receiver operating characteristic curve to assess imaging tests: a discussion and proposal for an alternative approach. Eur Radiol. 2015 Apr;25(4):932-9. PMID: 25599932. doi: 10.1007/s00330-014-3487-0.
47. van der Ploeg T, Austin PC, Steyerberg EW. Modern modelling techniques are data hungry: a

simulation study for predicting dichotomous endpoints. BMC Medical Research Methodology. 2014 2014/12/22;14(1):137. doi: 10.1186/1471-2288-14-137.

Supplementary Files

Figures

Articles' selection process.



Multimedia Appendixes

Search queries on Pubmed, Web of Science, SCOPUS and IEEE Xplore.
URL: http://asset.jmir.pub/assets/96c77cbb20a7a1369228777ba7608a31.doc

Characteristics of the included studies organized by model development stage, study design, setting and sample size: wound healing as outcome.
URL: http://asset.jmir.pub/assets/e53e7ac65b115aaa8c6131e25bfb8b13.doc

Distribution of clinical variables included in the final models by categories having healing as outcome.
URL: http://asset.jmir.pub/assets/ee7c9aa10abb3b3f33742b86be9ad4f3.doc

Characteristics of the included studies organized by model development stage, study design, setting and sample size: lower extremity amputation as outcome.
URL: http://asset.jmir.pub/assets/0acb37e4742c843fb4138cb77808dfb3.doc

Distribution of clinical variables included in the final models by categories having healing as outcome.
URL: http://asset.jmir.pub/assets/fb309cf2cff581c119430b8ef47e1a81.doc

Characteristics of the included studies organized by model development stage, study design, setting and sample size: mortality as outcome.
URL: http://asset.jmir.pub/assets/3b12ebe87a4243420176040631175e07.doc

Distribution of clinical variables included in the final models by categories having mortality as outcome.
URL: http://asset.jmir.pub/assets/acb45afe4372a30b5b64cf82e66588dc.doc

QUIPS tool results of included studies.
URL: http://asset.jmir.pub/assets/3bcf1338b2a19c49e531b095a7a3afdf.doc

PROBAST tool results of included studies.
URL: http://asset.jmir.pub/assets/eea3f04be0d8ef7a6bc020d9078d7f66.doc