Journal of King Saud University – Computer and Information Sciences 34 (2022) 3204–3225

Contents lists available at ScienceDirect

Journal of King Saud University – Computer and Information Sciences

journal homepage: www.sciencedirect.com

Machine learning and artificial intelligence based Diabetes Mellitus detection and self-management: A systematic review

Jyotismita Chaki [a,*], S. Thillai Ganesh [b], S.K Cidham [b], S. Ananda Theertan [b]

[a] School of Information Technology and Engineering, Vellore Institute of Technology, Vellore, India
[b] School of Electronics Engineering, Vellore Institute of Technology, Vellore, India

ARTICLE INFO

Article history:
Received 16 April 2020
Revised 29 June 2020
Accepted 30 June 2020
Available online 4 July 2020

Keywords:
Diabetes Mellitus
Machine learning
Artificial intelligence
Performance measure
Scopus database
PubMed database

ABSTRACT

Diabetes Mellitus (DM) is a condition induced by unregulated diabetes that may lead to multi-organ failure in patients. Thanks to advances in machine learning and artificial intelligence, which enables the early detection and diagnosis of DM through an automated process which is more advantageous than a manual diagnosis. Currently, many articles are published on automatic DM detection, diagnosis, and self-management via machine learning and artificial intelligence techniques. This review delivers an analysis of the detection, diagnosis, and self-management techniques of DM from six different facets viz., datasets of DM, pre-processing methods, feature extraction methods, machine learning-based identification, classification, and diagnosis of DM, artificial intelligence-based intelligent DM assistant and performance measures. It also discusses the conclusions of the previous study and the importance of the results of the study. Also, three current research issues in the field of DM detection and diagnosis and self-management and personalization are listed. After a thorough screening procedure, 107 main publications from the Scopus and PubMed repositories are chosen for this study. This review provides a detailed overview of DM detection and self-management techniques which may prove valuable to the community of scientists employed in the area of automatic DM detection and self-management.

Contents

* Corresponding author.
E-mail address: jyotismita.c@gmail.com (J. Chaki).

Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.jksuci.2020.06.013
1319-1578/

Nomenclature

Acronyms

DR	Diabetic Retinopathy
DT	Decision Trees
GA	Genetic Algorithms
PIDD	Pima Indians Diabetes Dataset
NB	Naïve Bayes
RF	Random Forest
ELC	Ensemble Learning Classifiers
CNN	Convolutional Neural Network
FFNN	Feedforward Neural Network
ELM	Extreme Learning Machine
LR	Linear Regression
BPNN	Back Propagation Neural Network
LFPN	Large-size Feature Pyramid Network
RCNN	Region Convolutional Neural Network
KNN	K-Nearest Neighbor
NN	Neural Network
SVM	Support Vector Machine
HRF	High-Resolution Fundus
LDA	Linear Discriminant Analysis
DL	Deep Learning
UC	Unsupervised Classifiers
ANN	Artificial Neural Network
MLP	Multilayer Perceptron
PRNN	Pattern Recognition Neural Networks
SVM-RBF	SVM with a Radial Basis Function kernel
FPN	Feature Pyramid Network
CF	Center Focus

1. Introduction

Important developments in biotechnology and, more importantly, high throughput computing is constantly contributing to quick and affordable data production, taking computational biology research into the world of big data. The primary goal is to investigate the increasingly growing biotic data and to build the framework for increasing responses to basic queries of medicine and biology. The efficiency and reliability of these techniques are obtained from the capacity of the correct approaches to identify forms and model creation from data. One of the most significant research applications is human-threatening disease prognosis and treatment. Diabetes Mellitus (DM) is one such disease (Artzi et al., 2020; Raman et al., 2016; Ramsingh and Bhuvaneswari, 2018).

DM is among the most widespread diseases (World Health Organization, 20202) for the elderly in the country. In 2017, 451 million individuals globally are diabetic as informed by the International Diabetes Federation. Expectations are that this figure will rise to 693 million citizens over the next 26 years. The primary cause of DM remains unclear, but researchers believe that both environmental and genetic factors play an important role in DM. While it is incurable, medications and drugs may be used to control it. Individuals with DM are at danger of having additional health complications, like cardiac arrest and organ damage. Early detection and management with DM will also avoid complications and help to decrease the threat with severe health issues.

Diagnosis of DM may be done either manually by a medical practitioner or by an automatic device. Any of these forms of measurement of DM involve benefits and drawbacks. The main advantage of manual diagnosis is that it does not need any help from the machine for the DM detection procedure, thus allowing the medical professional to be a specialist in the area. Often the symptoms of DM in its initial phase are so low that even an experienced doctor can't fully identify them. As a result of advances in Machine Learning (ML) and Artificial Intelligence (AI), the disease detection and diagnosis at an initial stage by an automated program is more probable and efficient than the manual DM recognition method (Sharma and Singh, 2018; Afzali and Yildiz, 2018; Theera-Umpon et al., 2019; Zou et al., 2018; Alghamdi et al., 2017). Benefits include a decreased workload for medical practitioners and a smaller risk of fault caused by humans. For efficient diagnosis and profitable management, decision support systems that are built on the computer may perform a vibrant role. DM field produces big data based on laboratory valuation, reports about the patient, treatment, follow-ups, medicine, etc. It is difficult to manually assemble all data appropriately. The quality of the organization of data has been affected because of unsuitable data management. Improvement in the data amount requires some suitable way to extract and process data efficiently and effectively. Machines for data collection and inspection are hired in modern and new hospitals to make them able for data collection and share in big information systems. An automated device is capable to identify DM and handle anomalies with far better simplicity and reliability compared to manual detection and diagnosis. Automation of the diagnosis of DM is therefore important. Automated DM systems may be built either by machine learning approaches or by artificial intelligence approaches.

All ML and AI methods have benefits and limitations of their own. Both methods have therefore been used to build automatic DM detection systems. The AI and ML-based techniques need causability and explainability to perform like a human (Holzinger et al., 2019) as many best-performing techniques of Ml and AI are least transparent. The explain ability of AI systems helps to improve the faith of doctors in future AI systems. The causability is based on the causal model which is measured in terms of efficiency, effectiveness related to causal understanding, and its transparency for a user. Several researchers have used ML and AI methods for DM control and self-management and personalization in recent years. Nevertheless, only a few review papers on DM detection and diagnosis procedures have been published (Gupta and Chhikara, 2018; Cruz-Vega et al., 2019; Choudhury and Gupta, 2019; Sun and Zhang, 2019; Spaggiari et al., 2018; Xiong et al., 2018).

The main contribution of this review article is the consideration of both ML and AI-based approaches in DM detection, diagnosis, self-management, and personalization. As we know, this is the first review article that covers both ML and AI for the detection, diagnosis, and self-management of DM and personalization of DM therapy. Review papers are relevant because they summarize the current research in a particular area in a detailed manner. Also, authors have only studied ML procedures, but certain essential facets of ML, like databases, pre-processing methods, and feature extraction and selection approaches used to identify DM and AI solutions to the need for intelligent DM assistants, are not addressed. As a consequence, attempts have been made in the sense of this analysis to examine existing literature on ML and AI approaches to DM studies.

Because of the variety and complexity of DM detection and diagnosis and self-management and personalization systems, a systematic decision-making framework is used for the selection of papers obtained from the Scopus and PubMed database. The purpose of this framework comprises of, (1) Datasets description, (2) Pre-processing techniques, (3) DM feature selection methods, (4) DM detection using ML approaches, (5) Intelligent DM assistant using AI methods and (6) performance matrices. After thorough exploration procedures, a total of 107 current related significant studies have been listed from the Scopus and PubMed databases. It is anticipated that this review will benefit the research communities in the field of DM detection and diagnosis and self-management and personalization discipline. The key analysis objectives of this review are:

1. Publicly accessible and self-created datasets in the area of DM detection.
2. Pre-processing methods that apply to DM datasets.
3. Widely utilized ML feature extraction techniques in the area of DM detection.
4. Widely utilized ML-based techniques for detection, classification, and diagnosis of DM.
5. Widely utilized AI-based techniques for intelligent DM assistant for self-management and personalization of DM therapy.
6. Performance matrices that are used to assess DM detection and diagnosis algorithms.
7. Future research directions for research to be resolved by future scientists working in the area of DM detection and diagnosis.

The review is organized in this fashion: Section 2 delivers a concise presentation of the DM disease, Section 3 provides different machine learning (ML), artificial intelligence (AI), and knowledge exploration techniques. Section 4 provides the research approaches utilized for the selection of the articles, Section 5 provides the review of datasets used by the researchers in the area of DM, Section 6, presents the pre-processing techniques, Section 7 presents feature selection and analysis, Section 8 provides the review of ML and AI techniques in the field of DM. This section includes a detailed review of ML techniques used in the detection and diagnosis, classification, and diagnosis of DM as well as the review of AI-based intelligent DM assistance and analysis of different performance measures. Section 9 is a discussion, followed by Section 10 which provides some Future Scopes and Challenges related to DM. Section 11 concludes the paper.

2. Diabetes Mellitus

Diabetes Mellitus (DM), sometimes called diabetes, is a concept for a variety of disorders that include how the body converts food into energy. Once one consumes food, the body converts it into sugar named glucose and transfers it to the bloodstream. The pancreas produces insulin, which is a hormone that tends to transfer glucose from the blood to the cells that utilize it for energy (Qureshi et al., 2019; Natarajan et al., 2019; Manikandan, 2019; Srividhya and Muthukumaravel, 2019).

If you have DM and don't seek medication, the body doesn't produce insulin as it does. Very much glucose persists in the body, a disorder commonly called high blood sugar. It may trigger severe or life-threatening health issues. DM develops in various ways, depending on the source (Carracedo et al., 2019).

2.1. Prediabetes

Prediabetes occurs whenever the blood sugar rises higher than it should be, but still not strong enough for the doctor to recognize diabetes. With prediabetes, the risk of type-II diabetes and cardiac disease will increase. Exercising further and reducing excess weight, sometimes less than 5%–7% of the body weight, will reduce these dangers (Bansal, 2015).

2.2. Type-I diabetes

Type-I diabetes is often denoted to as insulin-dependent DM. This also named juvenile-onset DM, as it frequently occurs in infancy. Diabetes with type-I is an autoimmune condition. This occurs as the body threatens the pancreas with antibodies. The organ is weakened and does not produce insulin (Katsarou et al., 2017). Body genes can cause this sort of diabetes. This may also happen due to complications with cells in the pancreas that produce insulin. Many of the health issues that might happen with the type-I result due to disruption to narrow blood vessels in the kidneys (diabetic nephropathy), eyes (termed diabetic retinopathy), and nerves (diabetic neuropathy). Anyone of type-I is often at greater risk of heart failure and stroke.

2.3. Type-II diabetes

Type-II diabetes has been known to be named non-insulin-dependent or adult-onset diabetes. Yet it has been prevalent in children and teenagers in the last 20 years, mainly as more young people are obese or overweight. Approximately 90% of patients have type-II diabetes (DeFronzo et al., 2015).

The pancreas normally releases some insulin while you have type-II diabetes. Yet either that isn't enough, or the body doesn't utilize it as it would. Type-II DM is also relatively mild than type-I diabetes. Yet it may also cause significant health problems, particularly in the small blood vessels in nerves, kidneys, and eyes. Type-II also raises the chances of stroke and heart failure.

Those who are overweight — more than 20% above their ideal body weight due to their height — have a very elevated chance of type-II diabetes and the health complications that may occur. Obesity also induces insulin resistance, and the pancreas has to function harder to produce more insulin.

2.4. Other forms of diabetes

The origin could be other factors in 1% to 5% of individuals who have diabetes. Those have pancreatic disorders, other operations and drugs, and illnesses. Under these situations, the doctor may want to keep an eye on the blood sugar levels (Flannick et al., 2016).

While comprehensive DM research has generated considerable information over the previous few decades on a) etiopathology (cellular mechanisms and environmental or genetic causes), b) diagnosis and c) disease detection, diagnosis, and control, more remains to be found, unfolded, explained and demarcated. In this attempt, depending on a huge, fast and increasingly growing body of clinical evidence and research helps to provide a substantial base for effective evaluation and follow-up. ML and AI, therefore, appear as core technologies, leading significantly to clinical decision-making. The goal is therefore to connect the assessment of the data with the treatment and the smart decision-making in the implementation and application of drugs.

3. ML, AI, and knowledge exploration

ML and AI is a research discipline that deals with the way computers learn from experience. To certain researchers, the phrase "ML" is the part of "AI," provided that the capacity to learn is the coarse attribute of an intellectual individual. The goal of machine learning is to develop computer systems that can learn and respond from their prior observation. The goal of artificial intelligence is to develop an intelligent agent or assistant that use different machine learning techniques-based solution (Al-Taee et al., 2016; Thompson and Baranowski, 2019; El-Sappagh et al., 2018; Sosale et al., 2020; Chatrati et al., 2020).

Knowledge exploration in databases (KEDs) is a discipline that incorporates hypotheses, approaches, and strategies, tries to understand the data and derive valuable facts from it. It is known to be a multi-step method (selection, pre-processing, transformation, ML/AI, understanding/assessment) defined in Fig. 1. The most critical phase in the whole KED method is ML/AI, which exemplifies the use of ML algorithms and AI in the data processing.

ML processes are usually categorized into three specific groups. These include 1) Supervised Learning (SL), where the scheme indicates the functionality of the labeled training data; 2) Unsupervised Learning (UL), where the system attempts to deduce the nature of unidentified data; and 3) Reinforcement Learning (RL), where the machine communicates with a dynamic context. Artificial intelligence is used to develop intelligent assistants which will help in self-management and personalization of the disease therapy.

3.1. Supervised learning

SL is where an algorithm is used to learn the mapping function from the input (I) to the output (O).

$$O = f(I)$$

The purpose is to determine the mapping function so accurately that the output variables (O) for that system can be predicted when a new input data (I) occurs (Taherkhani et al., 2018).

There are two categories of learning processes in SL: classification and regression. Classification methods aim to simulate distinct classes, like genotypes, whereas regression methods forecast real values. Some of the most common methods are KNN, DT, NN, GA, and SVM.

3.2. Unsupervised learning

UL is where there are only input data (I) with no associated output variables. The objective of UL is to simulate the basic nature or distribution of data to understand more about the content. Procedures are left to their devices to explore and display the fascinating content structure (Krotov and Hopfield, 2019).

There are two types of learning processes in supervised learning: association and clustering. The clustering is where the underlying groupings are revealed in the data, while the association rule learning method discovers the rules that define significant portions of the data.

3.3. Reinforcement Learning

RL is a generic term offered to a class of strategies through which the method aims to improve by active contact with the world to optimize any idea of incremental reward. It is necessary to remember that the program has no previous awareness of the actions of the world and the only way to realize is by trial and error (Silver et al., 2018). This type of learning is particularly applicable

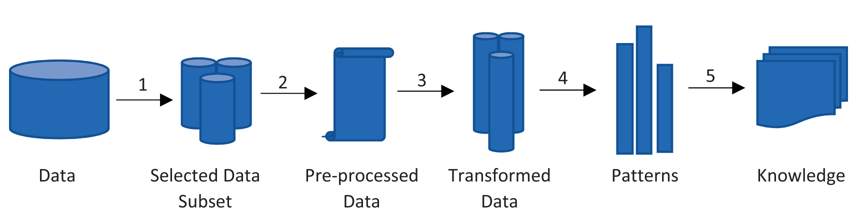

Fig. 1. The basic phases of KED: (1) Selection of data, (2) Pre-processing of data, (3) Transformation of data, (4) ML/AI, (5) Understanding/Assessment.

to autonomous devices, because of their flexibility about their environment.

4. Article selection process

Extensive attempts have been made to find papers utilizing machine learning and artificial intelligence methods for DM research. Two most popular databases i.e., Scopus and PubMed searched (08 June 2020) for a research article. The primary justification for the usage of these two databases is that there is a large collection of high-impact academic research papers in the area of the medical field and computer science.

There is a close connection between ML and AI. As a consequence, machine learning approaches are also referred to as artificial intelligence techniques in science literature. To resolve this and to be more precise in locating all relevant manuscripts, two searches are conducted in Scopus and PubMed, created on the following requests: a) "Diabetes" AND "Machine Learning" (DML), and b) "Diabetes" AND "Artificial Intelligence" (DAI).

Because of the large number of papers retrieved from these two queries (DML: 1000 and DAI: 1403), our scope of search is restricted to articles published during the last six years (2015–2020), which substantially narrowed the collection (DML: 300 and DAI: 450). It is necessary to say that a large number of publications is because papers are not only restricted to the domains of ML and AI, but rather to the wider area of computer science.

The next phase is a manual review of all retrieved documents. The main aim of this manual review is to first determine their contribution to DM research. Also, for DML, manual evaluation is carried out to remove papers that do not include machine learning techniques; for example, papers with basic statistical analyses. Finally, concerning DAI, the intention of the manual review is twofold. The first is to locate all papers relevant to machine learning and the second is to find articles related to an intelligent agent-based assistant for DM. Manual review narrowed the list (DML: 79, and DAI: 28), culminating in a total selection of 107 papers. These publications are categorized into the following two categories: Detection, Classification, and Diagnosis of DM; Intelligent DM Assistant. The entire collection method of the paper is shown in the workflow (Fig. 2).

5. Review of dataset used by researchers

In the chosen papers, publicly accessible datasets are used by most authors and others utilized unique self-created datasets that are divided into two sets: training and testing. For example, (Huang et al., 2018) used 16 images collected from MESSIDOR, 21 images from DiaRetDB0, and 5 images from DiaRetDB1 for the automated detection of neovascularization in retinal images. Images are selected automatically from databases that are classified as proliferative. (Swapna et al., 2018) collected Electrocardiograms (ECG) of 20 individuals each with DM and a typical group of people lying in a stable supine posture for 10 min. The data for the cardiac rate is extracted from ECG signals for DM detection.

Table 1 provides several diabetic databases used in selected research papers. The table lists the names of the datasets, a summary and URL of each of them, and the references of the publications in which the datasets are used.

Out of 107 studies, most of the time researchers used publicly obtainable datasets. PIDD 13 times; Messidor database 9 times; Kaggle DR dataset 6 times; DIARETDB0 Database 6 times; DIARETDB1 Database 6 times; STARE dataset 4 times; DRIVE Database 3 times, UCI machine learning DR dataset 2 times, EyePACS dataset 2 times; Review-DB, IDRiD dataset 2 times, and ICSM, HER, CPCSSN and HRF Image Database are used only one time. In some studies, researchers created the database as per their need (Self-created databases.) As a consequence, publicly accessible databases can be regarded as standard databases, since a variety of research and tests are performed on them.

6. Review of Pre-processing techniques

Diabetic data is pre-processed for a proper representation of the network to obtain more distinguishing characteristics until the data is clear. The foregoing is a short overview of the pre-processing methods utilized by researchers in selected scientific papers.

PIDD includes a variety of incomplete and impractical attributes, such as 0 plasma glucose and 0 body mass index. Several researchers (e.g., Chen et al., 2017) use data pre-processing by substituting mean for incomplete values and unlikely values.

The extraction of the green channel is used on the RGB image to obtain the green channel of the image individually because it gives further detail on the context image (Ghani et al., 2019).

The procedure of adaptive thresholding is utilized efficiently for segmenting the optical disk from the fundus image. As an example (Lokuarachchi et al., 2019) used adaptive thresholding to segment red lesion candidates from the image.

Discrete Wavelet Transform (DWT) is applied to the green channel of the fundus image to reduce computation and complexity

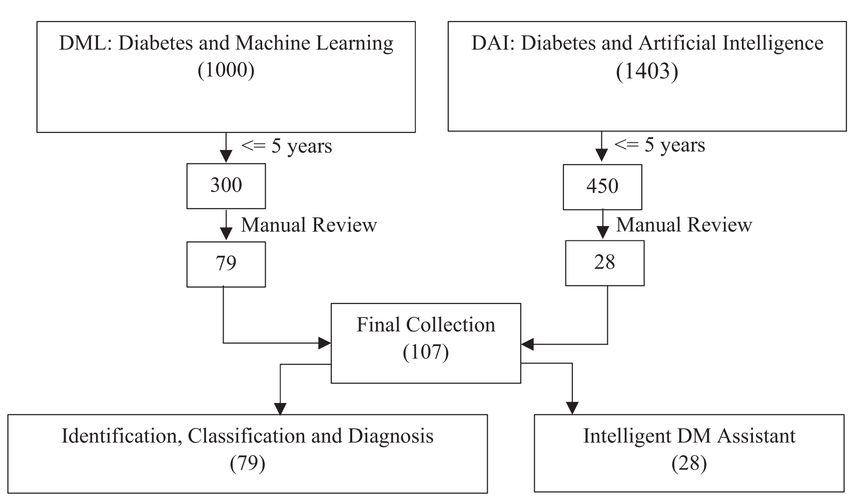

Fig. 2. Article selection process.

Table 1
Datasets Analysis.

Dataset	Description & URL	References
PIDD	The dataset comprises 768 females (both checked positive and negative for diabetes) of Pima Indian heritage living around Phoenix, Arizona with at least 21 years. URL: https://archive.ics.uci.edu/ml/datasets/diabetes	Chen et al. (2017), Xu and Wang (2019), Jahangir et al. (2017), Al-Zebari and Sengur (2019), Kriještorac et al. (2019), El-Sappagh et al. (2015), Sun et al. (2019), Shanthi et al. (2019), Sultan (2020), Jayashree and Kumar (2019), Benbelkacem and Atmani (2019) Nnamoko and Korkontzelos (2020), Bhuvaneswari and Manikandan (2018), Sneha and Gangil (2019), Gadekallu et al. (2020)
HRF image database	This dataset comprises three collections of retinal fundus images: healthy retinas, glaucoma images (focus and diffuse nerve fiber deficiency symptoms), and diabetic retinopathy (DR) (photos of patients with hemorrhoids, bright lesions, and spots). Every collection includes 15 images. URL: https://www5.cs.fau.de/research/data/fundus-images/	Ghani et al. (2019), Karthikeyan et al. (2019)
BioMedical Image and Signal Analysis (BIOMISA)	71 Optical Coherence Tomography and 71 fundus scans of 60 patients are there in the dataset. URL: http://biomisa.org/index.php/downloads/	Hassan et al. (2019)
Greek Food Composition Dataset	The following food categories are included in the dataset like (i) Egg product, (ii) Dairy product, (iii) Seafood product, (iv) Meat product, (v) Grain product, (vi) Oil, (vii) Vegetable product, (viii) Seed, nut, or kernel product, (ix) Sugar product, and (x) Fruit product. URL: http://www.eurofir.org/food-information/food-composition-databases/eurofir-aisbl-e-book-collection/	Konstantakopoulos et al. (2019)
Messidor dataset	This comprises of 1200 eye fundus color numerical images collected by 3 ophthalmological branches utilizing a 3CCD color video camera installed on a 45-degree non-mydriatic retinograph of the Topcon TRC NW6. Pictures are taken utilizing 8 bits/color plane at 2304×1536 or 1440×960, 2240×1488 pixels. URL: http://www.adcis.net/en/third-party/messidor/	Qomariah et al. (2019), Chetoui et al. (2018), Prabhu et al. (2019), Chandakkar et al. (2017), Huang et al. (2018), Chandakkar et al. (2017), Nijalingappa and Sandeep (2015), Voets et al. (2019), Singh and Gorantla (2020)
EyePACS, LLC	This dataset exceeded 5 million retinal images. URL: https://www.eyepacs.com/	Kwasigroch et al. (2018), Voets et al. (2019)
Kaggle DR dataset	Patient Ids are mentioned to Images accompanied by a right or left tag. The data given comprises of noise. 35,126 photos with 5-group tags (end-stage, severe, moderate, mild, and normal). Datasets comprise of color images that range in width and height from the low thousands and the low hundreds. URL: https://www.kaggle.com/c/diabetic-retinopathy-detection/data	Arora and Pandey (2019), Arcadu et al., 2019, SISODIA et al., 2017, Li et al., 2019, Wijesinghe et al. (2019)
Stare dataset	This dataset includes 20 color images of eye fundus shot with a Top Con TRV-50 fundus camera at 35° field of view. The image resolution is 700×605 pixels. URL: https://cecas.clemson.edu/~ahoover/stare/	Mule et al. (2018), Chandakkar et al. (2017), Karthikeyan et al. (2019)
IDRiD dataset	The fundus images in this dataset are collected by a specialist utilizing a Kowa VX-10 alpha optical fundus camera with a 50○ field of view, all focused on the macula. URL: https://idrid.grand-challenge.org/	Benzamin and Chakraborty (2018), Singh and Gorantla (2020)
Self-created iris dataset	Iris images are collected from 15 non-diabetic persons (10 female and 5 male) and 11 diabetic persons (5 female and 6 male). The age range of the diabetic person is around 20–58 years, with an average age of 48 years. Although the age range for a non-diabetic person is about 20–51 years with an average age of 28 years.	Aminah and Saputro (2019)
Self-created ECG dataset	ECG of 20 participants each from DM and the normal community is obtained by lying people in a comfortable supine posture for 10 min. The ECG pulse is measured at 500 Hz. 71 dataset (same amount for DM and normal people) is derived from the reported results. Every dataset comprises of 1000 samples.	Swapna et al. (2018)
Review-DB	This database (Retinal Vessel Image Collection for the Estimation of Widths Database) is utilized to promote the creation and analysis of vessel measuring algorithms and consists of 16 photographs of 193 vessel segments. URL: http://www.aldiri.info/Image%20Datasets/Review.aspx	
Local Dataset	The local dataset is focused on the compilation of real-time images of the fundus at the Sagar Nursing Home Hospital in Shivamogga. The file includes retinal photographs composed of normal and different signs of DR.	Nijalingappa and Sandeep (2015)
Self-created fundus images	The dataset includes 5,198 images where the images are digitized to 2136×3216, comprising 500 fundus images and showing all 5 levels of seriousness. There are 10 labels on the lesion.	Chen et al. (2019)
Self-created meal data	Validation and training sets are produced utilizing a four-day setup. This example involves 18 meals: 4 breakfasts (between 7:00 a.m. and 8:30 a.m.), 6 snacks (2 in the morning, 3 in the afternoon and 1 in the evening), 4 lunches (between 12:30 a.m. and 2:00p.m.) and 4 dinners (between 8:00p.m. and 8:30p.m.). Overall 1,800 meal data for 100patients are separated into two parts: one half for the validation set and one half for the training set.	Aiello et al. (2018)

(continued on next page)

Table 1 (*continued*)

Dataset	Description & URL	References
Self-created blood sample data	There are 297 female and 179 male participants in the research. Participants are involved in this analysis when their body mass index is 25.0 to 30 kg/m^2 and classified as obesity. The samples are categorized into overweight persons (n = 225) and healthy persons (n = 251), their body mass indexes are 26.17 ± 0.84 kg/m^2 and 21.79 ± 0.92 kg/m^2 accordingly. A total of 18 blood samples and 16 biochemical samples are used throughout this analysis.	Chen et al. (2015)
Self-created diabetic data	Data collection is conducted in a mobile research center based in rural areas and cities in the state of Brandenburg, Germany (between 2016 and 2019). Participants (40–70 years of age) are studied in mobile testing systems to detect MetS. Waist circumference, blood glucose, HDL cholesterol, triglycerides, body mass, and blood pressure are calculated.	Datta et al. (2019)
Drive Database	This dataset consists of 40 colored retinal photographs arbitrarily picked from the 400 DR patients in the Netherlands. The photos are divided into training and testing sets of 20 items in each group. Of these 40 images, 7 go to the mild non-proliferative DR class, while 33 have no retinopathy sign. The images are 768 × 584 pixels in resolution, 8 bits/color channel, have a field of view of roughly 540 pixels in diameter, and are compressed in JPEG format. URL: https://computervisiononline.com/dataset/1105138662	Sangeethaa and Maheswari (2018), Karkuzhali and Manimegalai (2019)
Diaretdb0 Database	This database comprises 130 photos of the color fundus, of which 20 are regular and 110 include DR signs (soft exudates, hard exudates, microaneurysms, hemorrhoid, and neovascularization). Photos are taken with a 50° field-of-view optical fundus camera with unspecified device settings. URL: https://www.it.lut.fi/project/imageret/diaretdb0/	Sangeethaa and Maheswari (2018), Chandakkar et al. (2017), Huang et al. (2018), Karthikeyan et al. (2019), Anggraeni and Wibawa (2019), Nijalingappa and Sandeep (2015)
Diaretdb1 Database	DiaRetDB1 comprises of 89 color fundus photos, of which 84 show at least moderate non-proliferative indications (Microaneurysms) of DR and 5 are taken as regular and don't show any indications of DR according to all experts who participated in the assessment. Photos are shot utilizing the same 50-degree optical fundus camera field of view with variable imaging settings. URL: https://www.it.lut.fi/project/imageret/diaretdb1/	Sangeethaa and Maheswari, 2018, Chandakkar et al., 2017, Huang et al. (2018), Lokuarachchi et al. (2019), Anggraeni and Wibawa (2019), Nijalingappa and Sandeep (2015), Cao et al. (2018)
Self-created physiological and demographic data	Physiological and social data is obtained from 52 individuals. 27 participants are infected with DM (13 with type-II and 14 with type-I diabetes) while the other 25 are non-diabetic participants. A questionnaire is used to gather personal details from all participants.	Yin et al. (2019)
Self-created diabetic data	The dataset consists of 338 (158 non-diabetic and 180 DM) participants of age from 1 to 25 years with a mean interval of 7 years. The collection of subjects is created on three criteria, viz.: average age, gender ratio, length of diabetic period, and standard deviation.	Samant and Agarwal (2018)
Self-created hemoglobin data	The dataset is compiled by E.Czeczek-Lewandowska from 451 samples of students aged between 6 and 18 under the supervision of a children's diabetic clinic at Rzeszow State Hospital in Poland in 2016. The dataset is split into two categories depending on the findings of the HbA1c glycated diabetes hemoglobin analyses that are interpreted from the patient's medical history issued by the parental permission diabetic clinic.	Czmil et al. (2019)
National Health and Nutrition Examination Survey (NHANES) dataset	The dataset is created by National centre for Health Statistics to check the nutritional and health status of U.S. people. It contains about 5000 data which includes the clinical test results and survey of physical examination. The survey data comprises dietary, demographic, socio-economic, and health-related questions. The clinical test consists of physical, medical, and physiological measurements of diabetic patients done by doctors. URL: http://www.cdc.gov/nchs/nhanes/	Dinh et al. (2019)
Self-created dataset	This dataset consists of 417 data divided into four groups: diabetic patients without hyperlipidemia and hypertension (139), diabetic patients with hypertension (133), diabetic patients with hyperlipidemia (70), and diabetic patients with hypertension and hyperlipidemia (75). The age range of the patients is from 40 to 75 years.	Hao et al. (2019)
S(elf-created dataset	This dataset consists of 115 urine samples collected from University Hospital Coventry & Warwickshire, UK. Out of 115 samples, 72 samples are collected from the type-II diabetic patient and the remaining samples are from healthy people as the negative control.	Martinez-Vernon et al. (2018)
Self-created dataset	This dataset consists of 30,244 (17731 female and 12,513 male) high-quality fundus images. The age range of the patient varies from 8 to 98.	Jiang et al., 2019
Self-created dataset	The dataset comprises 200 retinal images collected from diabetic patients varying in the brightness, quality, and color. The	Jaya et al. (2015)

Table 1 (*continued*)

Dataset	Description & URL	References
	dimension of the images is 2240 × 1488 with resolution 96 dpi. Out of 200 images, 75 images do not contain exudates.	
ICSM dataset	HbA1c, BMI, triglycerides, total cholesterol, smoking habit, age, gender data collected from 1000 type-II patients to create the dataset. URL: https://www.icsm.gov.au/harmonised-data-model-version-1–2000	Dagliati et al. (2018)
Self-created dataset	This dataset comprises physiological, biochemical and sequencing data collected from 30 nondiabetic and 20 type-II diabetic patients.	Hathaway et al. (2019)
Self-created dataset	This dataset comprises of 420 ECG data collected from 16 type-I DM patients.	Ling et al. (2016)
Self-created dataset	This dataset consists of fundus data collected from 536 selected patients. Out of 536 patients, 430 data are normal and remaining data are diabetic retinopathy	Tsao et al. (2018)
Self-created dataset	This dataset comprises of insulin concentrations and plasma glucose recorded before and after glucose intake collected from 1492 patients.	Abbas et al. (2019)
EHR dataset	This dataset comprises of 9948 diabetes patient's electronic health record. URL: https://data.world/datasets/ehr	Nguyen et al. (2019)
CPCSSN dataset	This dataset consists of about 6 million data. URL: http://cpcssn.ca/	Lai et al. (2019)

with decent performance. Daubechies wavelet (db2) with two vanishing moments is chosen for the fundus image decomposition. For example, (Anggraeni and Wibawa, 2019) used DWT conversion that functions on data vectors whose length is 2^n, transformed to arithmetically different vectors of the identical length.

Outliers are identified from DM datasets using an improved outlier-based class approach. For example, (Jahangir et al., 2017) used this approach to distinguish outliers dependent on the outlier element of the class. With this process, the top 10 outliers are identified from a training set created on the 12 closest neighbors and correlation-dependent distance estimation. Such outliers are eliminated from the training set. Such an outlier safe DM dataset is then utilized by the testing classifier.

Table 2 displays the widely used pre-processing approaches in chosen scientific studies.

7. Feature selection and analysis

Feature Selection is one of the key concepts of machine learning that has an immense effect on the efficiency of the algorithm. The DM data features that are utilized to train ML models have a big effect on results. Feature Selection will be the first and most significant step in the design of the model. There are several benefits of DM feature selection like (1) Decreases overfitting: less repetitive data means less ability to render noise-based decisions; (2) Increases accuracy: less inaccurate data means improved simulation accuracy; (3) Decreases processing time: fewer data points decrease algorithm complexity and training of the algorithms become faster.

There are two key specific methods to the DM feature selection method. The first is to create a filter-based method where the DM feature set is filtered out before the development of the model. The second method is to utilize an ML algorithm to test various subgroups of DM features and eventually pick the one which is producing the best recognition accuracy. This procedure will be utilized to construct a mathematical model in the end. Techniques in this group are considered as wrapper approaches since the resultant algorithm covers the entire selection process.

Various types of features such as shape features, color features, texture features, and text features are utilized in DM research. The following subsections include a short overview of the following features.

7.1. Shape features

These features comprise different shape-related features that are used in DM research. For example, the shape features utilized by Ghani et al. (2019) for computing the segmented optic disk area for accelerating the retinal fundus image classification.

7.2. Color features

These features are focused on the colors of the image. For example, Chandakkar et al. (2017) used color correlogram features for the detection and diagnosis of DR. The correlogram feature is modified to be shifts invariant, illumination invariant, and most significantly, the unique color spectrum invariant of retinopathy images.

7.3. Texture features

These features provide significant texture related information of the diabetic dataset. Some scientists utilize the probability distribution of pixels as a feature in the case of the detection of diabetic retinopathy (Somasundaram and Alli, 2017). Some researchers (Xu and Wang, 2019) used random and weighted feature selection technique from the DM data.

7.4. Text features

These features give significant information about the diabetic patient in the form of text. For instance, a patient with DM has significant characteristics for the disease caused, based on glucose level, age, heredity, as well as other factors, and these characteristics vary from person to person (Sneha and Gangil, 2019). Some researchers used ECG based features like slope, amplitude, and width (Swapna et al., 2018) and some have used gene expression data (Wang and Liu, 2017).

The researchers used various features types in the selected studies related to ML strategies, as well as shape, color, texture, and text-based features. The most commonly utilized features are text-based and texture-based. Individually, the text-based features are used in 28 articles out of 107 articles. Texture features are used in 13 articles and shape features are used in 3 articles in 107 collected articles. The combination of Shape, Texture, and color is

Table 2
The widely used pre-processing approaches in chosen scientific studies.

P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	References
✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	Chen et al., 2017; Xu and Wang, 2019; Kriještorac et al., 2020; Datta et al., 2019; Sun et al., 2019; Jayashree and Kumar, 2019; Dagliati et al., 2018
✗	✓	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	Ghani et al., 2019
✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	Jahangir et al., 2017
✗	✓	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	Hassan et al., 2019
✗	✓	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	Huanga et al., 2018
✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	Qomariah et al., 2019
✗	✗	✗	✗	✗	✗	✓	✓	✗	✗	✗	✗	Arora and Pandey, 2019
✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	Kanungo et al., 2017
✗	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✓	Lokuarachchi et al., 2019
✗	✓	✗	✓	✗	✓	✗	✗	✗	✗	✗	✗	Anggraeni and Wibawa, 2019
✗	✗	✗	✗	✗	✗	✗	✗	✓	✓	✓	✗	Aminah and Saputro, 2019
✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	Benzamin and Chakraborty, 2018
✗	✓	✓	✗	✗	✗	✓	✓	✗	✗	✓	✓	Mule et al., 2019
✗	✗	✗	✗	✗	✗	✓	✓	✓	✗	✗	✗	Kwasigroch et al., 2018
✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	Swapna et al., 2018
✗	✓	✗	✗	✗	✓	✗	✗	✗	✗	✓	✗	Sisodia et al., 2017
✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	Chetoui et al., 2018
✗	✓	✗	✗	✗	✗	✓	✗	✗	✗	✗	✓	Prabhu et al., 2019
✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✓	Karkuzhali and Manimegalai, 2019
✗	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✓	Nijalingappa and Sandeep, 2015
✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	Chen et al., 2019
✗	✓	✓	✗	✗	✓	✗	✗	✗	✗	✗	✓	Sangeethaa and Maheswari, 2018
✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	Samant and Agarwal, 2019
✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	Li et al., 2019
✗	✗	✗	✗	✗	✗	✗	✓	✓	✗	✗	✗	Yin et al., 2019
✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	Sultan, 2020
✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	Samant and Agarwal, 2018
✓	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	Dinh et al., 2019
✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	Martinez-Vernon et al., 2018
✗	✗	✗	✗	✗	✓	✓	✓	✓	✗	✓	✗	Jiang et al., 2019
✗	✗	✗	✗	✗	✓	✓	✓	✗	✗	✗	✗	Singh and Gorantla, 2020

**P1: Replace the missing values and irrelevant values; P2: Green Channel Extraction; P3: Adaptive Thresholding; P4: Discrete Wavelet Transform; P5: Outlier Detection; P6: Illumination Correction; P7: Resize; P8: Augmentation; P9: Data Normalization; P10: Grayscale Conversion; P11: ROI selection; P12: Histogram equalization

used 10 times; shape and texture are used 4 times; shape and color used 2 times; color and texture used 5 times and texture and text-based features used 5 times in 107 selected articles. From this detail, this can be inferred that text and texture features are the most discriminatory features of DM detection and diagnosis algorithms that can yield remarkable outcomes for scientists operating in the area of ML-based DM detection and diagnosis and recognition.

Table 3 summarizes 107 research papers with the adapted features.

8. Review of ML and AI techniques in the field of DM

The data-intensive aspect of DM diagnosis and treatment renders it perfect for incorporating artificial intelligence and machine learning to enhance results and identify innovative approaches. In the chosen primary studies of machine learning methods, the researchers used various kinds of machine learning algorithms that are used in various researches related to DM. Different types of intelligent assistants for DM developed by using artificial intelligence techniques are also reported in different studies. This por-

Table 3
Features adapted in selected researches.

Shape	Color	Texture	Text	References
✗	✗	✓	✗	Somasundaram and Alli, 2017; Xu and Wang, 2019; Mule et al., 2019; Lokuarachchi et al., 2019; Anggraeni and Wibawa, 2019; ALI et al., 2020; Aminah and Saputro, 2019; Chetoui et al., 2018; Nazir et al., 2019; Gadekallu et al., 2020; Nijalingappa and Sandeep, 2015; Aminah and Saputro et al., 2019; Li et al., 2019
✓	✗	✗	✗	Ghani et al., 2019; Prabhu et al., 2019; Lekha and Suchetha, 2017
✗	✗	✗	✓	Sneha and Gangil, 2019; Swapna et al., 2018; He et al., 2019; Nnamoko et al., 2020; Dai et al., 2018; Wang and Liu, 2017; Aiello et al., 2019; Kriještorac et al., 2020; Moreno et al., 2017; Chen et al., 2015; Datta et al., 2019; Bhuvaneswari and Manikandan, 2018; El-Sappagh et al., 2015; Sun et al., 2019; Yin et al., 2019; Shanthi et al., 2019; Severeyn et al., 2020; Sultan 2020; Jayashree and Kumar, 2019; Benbelkacem and Atmani, 2019; Czmil et al., 2019; Dinh et al., 2019; Hao et al., 2019; Martinez-Vernon et al., 2018; Dagliati et al., 2018; Hathaway et al., 2019; Ling et al., 2016; Tsao et al., 2018; Abbas et al., 2019; Nguyen et al., 2019, Lai et al., 2019.
✓	✗	✓	✗	Hassan et al., 2019; Benzamin and Chakraborty, 2018; Karkuzhali and Manimegalai, 2019; Samant and Ravinder, 2018
✓	✓	✗	✗	Konstantakopoulos et al., 2019; Huang et al., 2018
✗	✓	✓	✗	Qomariah et al., 2019; Chandakkar et al, 2017; Jaya et al., 2015; Voets et al., 2019; Singh and Gorantla, 2020; Cao et al., 2018.
✓	✓	✓	✗	Kwasigroch et al., 2018; Arora and Pandey, 2019; Arcadu et al., 2019; Kanungo et al., 2017; Karthikeyan et al., 2019; SISODIA et al., 2017; Chen et al., 2019; Sangeethaa and Maheswari, 2018; He et al., 2019
✗	✗	✓	✓	Samant and Agarwal, 2019; Jiang et al., 2019

tion is therefore intended to review the machine learning algorithms used in identified primary studies as well as different intelligent assistants utilized in DM research. In general, ten different ML algorithms are mainly utilized in specified primary studies. These are; KNN, SVM, NB, DT, LDA, RF, DL, ELC, Fuzzy based methods, and UC. Specifics of these algorithms are presented in the subsequent sections.

8.1. Detection, classification, and diagnosis of DM

In the chosen primary studies of machine learning methods, the researchers used various types of ML algorithms to create the DM detection and diagnosis and classification model.

8.1.1. K-nearest neighbor classifier

KNN algorithm assumes that similar DM data exist nearby. First, the classifier loads the DM training and testing data. Then the K value is selected by some investigational study to choose the number of neighbors. For every DM testing data, the distance is calculated for each training data point. The calculated distances are stored from smallest to largest by the distances. After that, pick the first K entries. The DM test data then allocated to the class based on most of the classes exist in the selected points.

Some researchers, like Nijalingappa and Sandeep (2015), Aminah and Saputro (2019); Carter et al. (2019), Chandakkar et al. (2017), Lokuarachchi et al. (2019), have used KNN algorithms to identify various diabetic types.

Ali et al. (2020) utilized various types of KNN algorithm to identify and classify DM. The database is created on the specifications of the American Diabetes Association. For the training level, 4900 samples are utilized by the learner classifier method to analyze the output. Thereafter, 100 of the data samples are utilized for the test. The findings indicate that the KNN types (Fine, Medium, Weighted, and Cubic) have strong precision over the Coarse and Cosine methods. Fine KNN is known to be the most relevant due to its precision of classified samples.

The iridology-based or iris-based DM prediction framework is developed utilizing ML algorithms (Aminah and Saputro, 2019). ML is utilized to automate the detection process. The built framework comprises of methods for the creation of eye images and algorithms for image processing. Iris images are shot utilizing the Iriscope Iris Analyzer Iridology system. The Gray Level Co-Occurrence Matrix (GLCM) approach is utilized for the feature extraction to acquire the texture characteristics of the image. The KNN approach is utilized to distinguish non-diabetic and diabetic groups. Categorization tests are then checked utilizing the k-fold cross-validation process and analyzed utilizing the confusion matrix. Two classes of objects are assessed: one is 16 non-diabetic subjects and 11 are diabetic subjects. The findings reveal that the precision is 85.6% the false-positive rate (FPR) is 11.07%, the false-negative rate (FNR) is 20.40%, the specificity is 0.889, and the sensitivity is 0.796.

Type-I DM is a chronic metabolic condition induced by cell-mediated autoimmune degradation of pancreatic beta cells, so exogenic insulin management is required to control glycemia. The method suggested by (Aiello et al., 2018) uses several KNN classification methods to estimate the postprandial glucose profile owing to minimal treatment and to recommend time and/or meal bolus adjustment. This method has been broadly tested for adults in the silicon community of the UVA/PADOVA system and has been approved by the Food and Drug Administration as a supplement for animal research.

8.1.2. Naïve Bayes

NB counts the feature value occurrences and unions of feature values in DM data with the persistence of utilizing probabilities. It calculates explicit probabilities for hypothesis and is also sensitive to noise in DM data. If N is the number of training data, and T of them are in type t, then the probability of type e is:

$$P(t) = T/N' \tag{1}$$

For DM data, S in the testing set, there are G samples $G_S = \{G_1, G_2, \ldots, G_X\}$. The purpose is to classify this DM data into t category. The probability that S goes to t category is:

$$P(t|S) = \frac{P(S/t) \cdot P(t')}{P(S)} \prod_{G=P\left(\frac{G_S}{t}\right) \cdot P(t)}^{G} \quad (2)$$

Some researchers, like Sparapani et al. (2019), Aminah and Saputro (2019), have used NB to identify various diabetic types.

The approach suggested by Sneha and Gangil (2019) is intended to concentrate on the selection of characteristics correlated with early detection of DM using a predictive model. They used a sample DM dataset (2500 data items) with 15 attributes obtained from the UCI machine repository. Experimentation is done using only 11 features selected by the proposed method. The outcome reveals that the NB test indicates 82.30% as the highest accuracy. Research often simplifies the selection of suitable sample attributes to boost the precision of the classification.

Patient degree of risk of DM has been observed using ontology and machine learning approaches (Lakshmi et al., 2019). The NB algorithm utilized to make choices on the medical record also determines the danger level possibilities. The proposed algorithm is tested against the following parameters, including the confusion matrix, the precision point, the mean, and this proposed study is considered to provide a better degree of accuracy relative to the current research.

8.1.3. Support Vector Machine

The utilization of SVM in DM data classification can be categorized as non-linear along with linear data. It uses non-linear mapping for altering DM training data into a higher dimension. After the alteration of the training data, it hunts for the linear ideal separating hyperplane. SVM attempts to split the DM data space with the use of non-linear or linear differentiation between different classes. Thus, in the SVM classifier, the perfect margins are created between different classifiers and then utilized for categorization to improve the training and testing speed.

Some researchers, like Sah and Sarma (2018), Sisodia et al. (2017), Chetoui et al. (2018), Karkuzhali and Manimegalai (2019), Mule et al. (2018), Lokuarachchi et al. (2019), He et al. (2019); Aminah and Saputro (2019); Carter et al., 2019)), have used SVM algorithms to identify various diabetic types. The authors recorded better classification efficiency using the SVM.

Qomariah et al. (2019) suggest an approach for the extraction of features and classification of DR using SVM. Authors use the high-level features of the last fully connected layer focused on transfer learning from the CNN as input features for classification utilizing the SVM. The suggested procedure is tested using 70 and 77 retinal photos from the Messidor database of Base 13 and Base 12. The maximum accuracy levels are 95.24% and 95.83% for base 13 and base 12, correspondingly, from the findings of the tests.

Wang and Liu (2017) developed a network system for gene co-expression to classify biomarkers with very specific trends of gene co-expression in DM disease. Authors categorize different genes into four classes through their pair-wise gene co-expression, i.e., gene pairs with high correlation in disease samples, without (and/or with) high control samples correlation. Then the primary genes are established as being strongly involved with DM pathogenesis. The classification capacity of such gene classes is assessed by a process based on the support vector machine. Genes with strong recognition specificity are further explored in the functional improvement study, which is known as biomarkers.

Heart rate variability (HRV) signals (obtained from ECG signals) may be utilized successfully for non-invasive DM detection. Swapna et al. (2018) provides a technique for classifying DM and regular HRV signals utilizing deep learning architectures. The authors used SVM for classification. The suggested classification method will enable physicians to detect DM by utilizing ECG indications with a very good precision of 95.7%.

8.1.4. Linear Discriminant analysis

LDA is a widely employed classification and dimensional reduction method. It may be used for multiclass classification. LDA builds on a line that necessarily retains directions that are valuable for the classification of results. It requires a line projection such that groups from various classes are segregated.

Moreno et al. (2016) propose a methodology of screening for the occurrence of type-II diabetes using a signal acquired from the pulse oximeter. The device comprises two portions: the first one examines the signal acquired from the device and the second one comprises an ML component. The device comprises a front end that derives a series of features of the device signal. Such features are created on biological criteria. The features collection is the input of an ML algorithm that defined the class of the test sample, i.e., if or not the person has DM. ML algorithms are gradient production, random forests, and LDA as benchmarks are used. The process is checked in a collection of 1157 samples (two tests per subject) obtained from five public health centers. The mean operational feature of the receiver is found to be with a precision of 64% for the threshold offering a sensitivity of 65%.

Carter et al. (2019) identified the usage of simple diabetic toenail analysis and ML strategies for the rigorous diagnosis of type-II DM. Cs, Aluminum, V, Ni, and Zn amounts in toenails are shown to be slightly different between healthy participants and type-II DM patients. Seven specific ML algorithms are then analyzed to create a non-invasive diagnostic approach utilizing toenail concentrations of twenty-two elements and personal details like gender, age, and smoking background as features. The LDA is being used as a classifier.

8.1.5. Decision Trees

A DT is derived from the DM training dataset, and the tree is then used in the future to assign the type characteristic attribute for the occurrences in the evaluation dataset. DT creation depends on heuristics, as distinct heuristics produce separate DTs from the same dataset. DTs identify the research samples based on their characteristics. Each non-leaf node in the DT is a component, and each branch is a value that can be obtained by the component. Samples are classified by following a route that begins at the root node and finishes at the leaf by following divisions, based on the characteristics of the diabetes data set. The value of the leaf defines the predicted value of the sample set.

Some researchers, like Sun et al. (2019), Lokuarachchi et al. (2019), Woldaregay et al. (2019), have used DT to identify various diabetic types.

Chen et al. (2017) suggest a hybrid forecast model to better detect type-II DM. In the proposed model, J48 decision tree is utilized as a classifier for recognition. To get an initial test, they used the PIDD dataset. The 90.4% outcome indicates that the proposed model is much more reliable than the other prior studies listed in the literature.

Al-Zebari and Sengur (2019) suggested an early-stage approach for the diagnosis of type-II diabetes to administer the appropriate medication. The PIDD dataset used by this study is accessible in the UCI machine learning repository. The DTs utilized in the assessment is the coarse tree (CT), the medium tree (MT), and the fine tree (FT) correspondingly. In the CT, MT, and FT, the highest number of splits is allocated to 4, 20, and 100, correspondingly. For DT approaches, the CT provided a precision score of 75.3%, which is the best of all DT approaches.

8.1.6. Random forest

RF is one of the most common and efficient classifying ML algorithms. It produces forest with DT. Usually, with more trees in the forest, better prediction can be achieved. To define a new model depending on the characteristics, every tree offers an identification vote and the tree label saves the model. The forest selects the party with the highest number of ballots. In other words, the RF classification process is close to the baggage strategy. A subgroup of the training set is generated in RF and a DT is built for every subgroup. When all DTs define each input vector for the test collection, the forest eventually selects the one with the most votes.

Some researchers, like Aminah and Saputro (2019), Carter et al. (2019), Benbelkacem and Atmani (2019), Samant and Agarwal (2018), Doreswamy and Santosh (2018), have used RF to identify various diabetic types.

Xu and Wang (2019) suggest a probability prediction model for type-II DM focused on the Ensemble Learning Process. In this model, the random forest-based weighted feature selection algorithm (RF-WFS) is used for the optimum collection of features and the XGBoost classifier for classification which is validated by using the PIDD DM dataset. The precision, specificity, and sensitivity of the model is 93.75%, 94.8%, and 91.79% respectively.

Prabhu et al. (2019) suggests an adaptive DR detection and diagnosis method focused on the existence of light retinal lesions, which is one of the signs of DR. Firstly, the optic disk is separated from the fundus image because its light is close to that of the bright lesions. Exudates are extracted and their various properties are attained. Later, a feature-based diagnosis is done to identify the various phases of the disease. Random Forest algorithm is examined and precision, specificity, and sensitivity are tested at each point of categorization. The accuracy attained utilizing RF is 100%, 85.71%, and 87.5% for Stage 1, 2, and 3 correspondingly.

8.1.7. Deep learning

DL methods extract features data and learn indirect features when trained that can differentiate between huge volumes of DM data. Also, if there are adequate numbers of samples and features of various variables are affecting the classification, then, DL methods may learn from these features easily. This can support when dealing with huge intraclass differences and noisy information be present in DM data. DL's capability to learn patterns from unrecognized DM information can improve social media information.

NNs are a non-linear method and are detected as a widely used DL model for the detection and diagnosis of DM. NN comprises of a huge number of extremely inter-reliant processing elements called neurons, employed together to solve any clear issue related to DM data. As they can extract considerable information from an enormous group of data, neurons are built for specific applications related to DM detection. Some DL models which are utilized in DM prediction are MLP (Mohebbi et al., 2017), FFNN (Ghani et al., 2019; Sultan, 2020), Deep CNN (Mohebbi et al., 2017; Kwasigroch et al., 2018; Arora and Pandey, 2019; Kanungo et al., 2017; Benzamin and Chakraborty, 2018; Karthikeyan et al., 2019; Chen et al., 2019; Lekha and Suchetha, 2017; Sangeethaa and Maheswari, 2018; Li et al., 2019), ELM (Anggraeni and Wibawa, 2019; Nazir et al, 2019; Chen et al., 2015; Huang et al., 2018; Shanthi et al., 2019); ANN (Prabhu et al., 2019; Raman et al., 2016; Dai et al., 2018; Konstantakopoulos et al., 2019; Yin et al., 2019), PRNN (Karkuzhali and Manimegalai, 2019; Jayashree and Kumar, 2019), Deep NN (Gadekallu et al., 2020), LR (Arcadu et al., 2019), etc. AutoMLP is auto-tunable and automatically conducts parameter optimization during the training phase, which otherwise needs human input for DM detection and diagnosis (Jahangir et al., 2017).

8.1.8. Ensemble learning classifiers

In ensemble learning several classification algorithms converge to create a more effective architecture. This can be achieved in two forms, i.e., baggage and boosting. In the bagging method, many classification procedures operate in parallel and eventually vote for the more reliable one. The entity with the largest vote is the ultimate classifier. Different classification algorithms are used in sequence for boosting. Each variable's weights are set based on the prior iteration. Next, the data is split into several sections, then one is checked with others, and so forth.

A few researchers like (Aminah and Saputro (2019); Somasundaram and Alli, 2017) have implemented ensemble classifiers in the selected papers.

Datta et al. (2019) propose an ML system for the initial diagnosis of DM. The model is trained and validated based on data obtained from a German community of 2314 subjects (female = 1396, male = 918). Of 2314 subjects, 941 have metabolic syndrome (female = 500, male = 441). Features included involve various anthropometric characteristics (such as weight, height, waist perimeter), drugs, gender, age, etc.; ML methods used included random forest, logistic regression, gradient boosters, and an ensemble model. They have a field under the curve values of up to 0.90 for the ensemble classifier.

Samant and Agarwal (2019) provide a thorough comparative study of the machine learning recognition approaches used to diagnose type-II diabetes utilizing a mix of physiological parameters and iris-based features. A selection of 334 topics is studied, separated into DM and non-diabetic classes. The diabetic group is split into three separate subgroups due to the length of the DM condition. Three kinds of ensemble classifiers are used, namely, the bagged tree (EBaT), boosted tree (EBoT), and subspace KNN (SKNN), and 85% precision is achieved.

8.1.9. Fuzzy-based methods

In anticipation of DM, Bhuvaneswari and Manikandan (2018) propose a new diabetic diagnostic method that incorporates a recently introduced temporal feature selection and a temporal fuzzy ant miner tree (TFAMT) classifier for successful recognition in the type-II diabetes study. Also, a new time-weighted genetic algorithm is mentioned in this research to boost recognition accuracy by pre-processing image and text data. Also, intelligent fuzzy rules are derived from the weighted temporal capability of TFAMT, and then the fuzzy rule extractor is utilized to minimize the number of functions in the derived regulations. The effectiveness of the suggested TFAMT – TWGA paradigm is tested utilizing the UCI registry and the DR dataset and achieves 83.7% precision.

El-Sappagh et al. (2015) propose a fuzzy ontology-based case-based justification paradigm for the treatment of DM. It introduces a fuzzy semantic retrieval algorithm and a fuzzy case-base OWL2 ontology that manages several forms of functionality. There are 60 possible diabetic instances in the fuzzy ontology. The resulting program will respond to complicated DM related questions linked to the abstract interpretation of medical principles and the handling of ambiguous terminology. The corresponding fuzzy case-base ontology includes 63 definitions, 54 (fuzzy) entity characteristics, 138 (fuzzy) datatype characteristics, 105 fuzzy data types, and 2640 instances. The device has a precision of 97.67%.

A novel fuzzy rule miner (ANT FDCSM) developed from an ant colony *meta*-heuristic for DM patients is suggested (Singh and Gupta, 2019) for the recognition of DM. A hybrid node split measure (SW FDCSM) is introduced to compute heuristic knowledge. SW FDCSM is a mixture of characteristic importance weight (SW) and a modern fuzzy version (Fuzzy DCSM) of the traditional separate class split test (DCSM). Enhancements are introduced to produce a detailed collection of laws while retaining reasonable precision (87.7%), sensitivity (92.2%), and specificity (80.3%). 10-fold cross-validation (10-FNo) is added to the PIDD data collection to verify the efficiency of ANT FDCSM.

Rahim et al. (2015) proposed a fuzzy system to detect DR. In the preprocessing part they have used fuzzy filtering, fuzzy histogram, and fuzzy edge detection from the retinopathy images. DT and

KNN classifiers are used for the classification of retinopathy images. The obtained accuracy is 74.6%.

8.1.10. Unsupervised classifiers

Typically, where relevant knowledge is not accessible, an unsupervised description is utilized. In this case, only the data collection and the attributes that apply to those cases are identified. Such a technique is used in the unsupervised grouping to identify clusters of points in a function space as is usually achieved for clustering methods.

UC algorithms have been used many times in the selected studies. Severeyn et al. (2019) used the k-means clustering algorithm to classify samples with type-II DM and pre-diabetes utilizing the glucose and insulin. A database of 188 samples with insulin and glucose values is used during the oral glucose tolerance test. The k-means clustering obtained for glucose is above 0.6 (silhouette coefficient) in all cases.

Table 4 demonstrates the research-wise use of ML algorithms in each chosen main analysis. In Table 4, the 1st column displays the ML algorithms utilized by various scientists in the 79 selected articles, the 2nd column displays the best algorithm that exceeds the performance of other algorithms and achieved the maximum classification performance. The 3rd column indicates the algorithms where the primary methodology is compared and the efficiency of the best methodology is finally shown. There are few experiments in which scientists have utilized just one method and have not compared their ML classification strategies to any other classification method. Thus, the 3rd column value is blank (~) in these instances. Various evaluations have been used in multiple experiments for classification purposes, and the effects of the collected values have been seen with correct performance measures.

8.2. Intelligent DM assistant for self-management and personalization of DM therapy

Rapid developments in AI aim to render both standardized and unstructured real-time health data accessible for DM treatment (Burki, 2015) and personalization of DM therapy (Donsa et al., 2015). The Turing Library for Computer History describes AI as "the science of having machines do things that need knowledge as performed by humans". AI covers a wide range of methods to simulating human intelligence and performing various thinking tasks in the area of DM research.

Wijesinghe et al. (2019) suggest a project involving an autonomous device named the Intelligent Diabetes Assistant (IDA) that determines the diagnosis and care prioritization based on the findings seen on the computer. The IDA comprises of knowledge-based applications for seriousness level assessment, medical conclusion assistance, and near-real-time foot ulcer diagnosis and border monitoring. The System Usability Scale (SUS) is utilized to assess the usefulness of the IDA in terms of efficiency, learning, and satisfaction.

Maharjan et al. (2019) suggest an AI-powered voice-based virtual assistant (Alexa) that assist Native American DM patients to maintain their everyday food and gain health and nutrition-related skills. Voice is one of the most common forms of contact and is simple to utilize without any technological context. Alexa has been built on the platform of Amazon Alexa. The messages and advice generated by Alexa are customized based on the social, physical, and cultural profile of each individual. It will then be simple to be embraced by the target market. In this chat, the consumer queries Alexa to count his/her calories and also to obtain knowledge on health and nutrition awareness from Alexa.

Czmil et al. (2019) address a non-invasive type-I DM diagnosis system for children dependent on the assessment of physical activity. The key metrics for disease monitoring are weekly step counts and weekly minutes of vigorous activity. Dependence between the weekly steps number and the prevalence of type-I DM is developed after a detailed review of the data utilizing clustering and classification algorithms. The studies have provided positive evidence that type-I DM may be detected by evaluating physical movement. This is important concerning the non-invasiveness and versatility of the detection process, which can be checked anytime at any time. The current system may be applied on a mobile phone. The most acceptable findings are achieved utilizing the random forest approach.

Electronic noses (E-noses) may have a possible function to play in the screening/analysis of different respiratory and systemic diseases through the examination of respiratory signatures (Behera et al., 2019). E-nose combines a sensor array and an ANN that reacts to different volatile organic compound patterns and may also serve as a non-invasive disease detection system. Breath includes markers for DM together with markers for a variety of diseases and, therefore, a non-invasive methodology like E-nose will significantly enhance diagnostic techniques over current invasive approaches.

Al-Taee et al. (2016) introduce a modern eHealth framework that incorporates humanoid robots to promote an evolving multi-dimensional method to DM care. The network design applies the IoT to a cloud-centric framework by leveraging current domain protocols to connect and manage physical layer objects. This involves capillary networks, each of them contains a medical sensors series connected to a humanoid robot wirelessly (through the Internet) to a web-based disease control center. A completely usable system is being built and its end-to-end design and appropriateness are being actively validated in a clinically-led pilot trial, which offers proof that both patients and caregivers are responsive to the new model.

Vaskovsky and Chvanova (2019) propose a personalized food recommendation system (Ramus) which can be used to recommend nutrition and food for diabetic patients. To develop this author used an electronic medical report and stored patient's taste preferences. Then the system is trained with the data which includes various substances and food effects on patients. After training the system with these data, food recommendation output can be formed.

Demongeot et al. (2015) used a game-theoretic approach to personalize the food habits of diabetic patients. The food recommendation system gives several menus as input and the virtual advisor reminds the food mistakes of the patient and advises the food the patient should take to his diabetic condition. This system also recommends different exercises the patient should adapt throughout the year.

Table 5 provides a review of intelligent DM assistant used by the researchers collected from 28 articles.

8.3. Review and analysis of different performance matrices

Table 6 displays the different performance metrics utilized in the chosen research papers. A detailed description of performance metrics comprises of accuracy, specificity, sensitivity, the area under the curve (AUC), precision, F-score, and positive predictive value, the negative predictive value can be seen in Sokolova and Lapalme (2009). Likewise, the discussion on PABAK, Kappa Score can be observed in Chen et al. (2009). Throughout the majority of chosen scientific papers, the researchers used the specificity, accuracy, sensitivity, and area under the Receiver's Operational Curve (AUC) as their measurement criteria to assess the efficiency of the classifier.

** Acc: Accuracy, SEN: Sensitivity, SPE: Specificity, AUC: Area Under Curve, FSc: F-Score, Prec: Precision, NPV: Negative Predictive Value, KSc: Kappa Score, PPV: Positive Predictive Value.

Table 4
Review of Machine Learning classifiers in the field of DM.

References	Field of use	Classifier	Compared with	Performance
Mohebbi et al. (2017)	Detection of type-II diabetics	CNN	LR, and MLP	77.5 ± 1.4% AUC
Chen et al. (2017)	Detection of type-II diabetics	DT	ANFIS, RF, KNN	90.04% Accuracy
Somasundaram and Alli (2017)	DR	ELC	Automated Micro Aneurysm Detection, Age-related Macular Degeneration - DR, Secure Identity Recognition Scheme based on Electroencephalogram, and Multi-Scale NonNegative Sparse Coding	85% Accuracy
Xu and Wang, 2019)	A Risk Prediction Model for Type-II Diabetes	XGBoost	C4.5, NB, AdaBoost, RF	93.75% Accuracy
Ghani et al. (2019)	Retinal Fundus Image Classification	FFNN	Circular Hough transform, Blood Vessel convergence illumination equalization, Mean, standard deviation and cup-disk ratio	100% Accuracy
Jahangir et al. (2017)	DM prediction	AutoMLP	Hierarchal Majority Voting based ensemble technique, C4.5, Neuro-Fuzzy Inference System, NB, Complex-Valued Neural Network	88.7% Accuracy
Sneha and Gangil (2019)	Early prediction of DM	SVM	RF, NB, DT, KNN	77.73% Accuracy
Hassan et al. (2019)	Retinal edema detection	Supervised discriminant analysis	~	93.33% Accuracy
Huang et al. (2018)	Detection of neovascularization in retinal images	ELM	SVM	78.7% Accuracy
Sah and Sarma (2018)	Bloodless Technique to Detect DM	SVM	~	91.67% Accuracy
Qomariah et al. (2019)	Classification of DR and Normal Retinal Images	Resnet 50 + SVM	Alexnet, VGGNet Type 16, VGGNet Type 19, Inception Resnet V2, Inception V3, GoogleNet, DenseNet201	95.83% Accuracy
Kwasigroch et al. (2018)	Detection and diagnosis and assessing the stage of DR	CNN	~	81.7% Accuracy
Arora and Pandey (2019)	DR Detection	CNN	~	74.4% Accuracy
Arcadu et al. (2019)	Measures of diabetic macular thickening	CNN	~	97% Accuracy
Kanungo et al. (2017)	Detecting DR	CNN	~	88% Accuracy
Mule et al. (2018)	Non-proliferative DR	SVM	~	95.3% Accuracy
Benzamin and Chakraborty (2018)	Recognition of Hard Exudates in Retinal Fundus Images	CNN	~	96.6% Accuracy
Karthikeyan et al. (2019)	Multi-Class Retinal Diseases	CNN	~	92% Accuracy
Lokuarachchi et al. (2019)	Recognition of Red Lesions in Retinal Images	KNN	SVM, DT, ELC	92.31% Sensitivity, 91.89% Specificity
Anggraeni and Wibawa (2019)	Recognition of the Emergence of Exudate on the Retina Image	ELM	~	65% Accuracy
Ali et al. (2020)	DM Classification	Fine KNN	Coarse KNN, Medium KNN, Cubic KNN, Cosine KNN, Weighted KNN	99.9% Accuracy
Aminah and Saputro (2019)	DM Prediction	KNN	SVM	85.6% Accuracy
Swapna et al. (2018)	DM detection	CNN-LSTM with SVM	CNN with SVM	95.7% Accuracy
Chetoui et al. (2018)	Retinopathy	Local Energy-based Shape Histogram with SVM-RBF	Local Binary Patterns with SVM-RBF, Local Ternary pattern with SVM-RBF	90.4% Accuracy
Nazir et al. (2019)	DR	Tetragonal local octa pattern with an ELM	Local Binary Patterns with extreme learning machine, LDP with extreme learning, Local Tetra pattern with extreme learning	99.5% Accuracy
Prabhu et al. (2019)	DR	RF	ANN	100% Accuracy
Karkuzhali and Manimegalai (2019)		SVM-RBF	PRNN	98% Accuracy
He et al. (2019)	Type-I Diabetes	Structured Output SVM	Linear SVM, NB, and SVM-RBF	84% Accuracy
Gadekallu et al. (2020)	DR	Deep NN	SVM, KNN, NB	96% Accuracy
Nnamoko and Korkontzelos (2020)	DM prediction using outliers	C4.5 DT	SVM-RBF, NB, Repeated Incremental Pruning to Produce Error Reduction,	89.5% Accuracy
Wang and Liu (2017)	Recognizing Biomarkers of Diabetes with Gene Coexpression Networks	SVM	~	96% area under the ROC curve

(continued on next page)

Table 4 (*continued*)

References	Field of use	Classifier	Compared with	Performance
Dai et al. (2018)	Blood Glucose detection	PSO-ANN	~	0.69 RMSE
Nijalingappa and Sandeep (2015)	DR	KNN	~	95% Accuracy
Chen et al. (2019)	Mini Lesions Detection on DR	LFPN + CF	FPN, LFPN, FPN + CF, FasterRCNN, FasterRCNN + CF	93.01% Sensitivities
Chandakkar et al. (2017)	Detection of DR	MIRank-KNN	~	96.55% Accuracy
Al-Zebari and Sengur (2019)	DM Detection	LR	DT, SVM, KNN, ELM, Discriminant Analysis	77.9% Accuracy
Aiello et al. (2018)	Meal Classification in Type-I Diabetes	KNN	~	36.15 RMSE
Lekha and Suchetha (2017)	Detection and classification of DM	CNN	SVM and NN	96.5% area under the ROC curve
Kriještorac et al. (2019)	Diabetes Mellitus type-II	Logistic Regression	DT, SVM, KNN	77.90% Accuracy
Moreno et al. (2016)	Type-II Diabetes Screening Test	RF	Gradient boosting, LDA	78.15% area under the ROC curve
Chen et al. (2015)	Identify the overweight status among DM people	Extreme learning	SVM with RBF kernel and BPNN	90.54% Accuracy
Dinh et al. (2019)	DM prediction	eXtreme Gradient Boost	Logistic Regression, SVM, RF, Weighted Ensemble Model	95.7% area under the ROC curve
Hao et al (2019)	Diagnose and monitor type-II DM	SVM with polynomial kernel	RF, LDA, Logistic Regression, SVM with linear kernel	96.3% Accuracy
Martinez-Vernon et al. (2018)	DM diagnosis	Sparse Logistic Regression	RF, SVM, Gaussian Process, NN	89.9% Accuracy
Jiang et al. (2019)	DR detection	Adaboost	Inception V3, Resnet152, Inception-Resnet-V2	94.6% area under the ROC curve
Jaya et al. (2015)	Fundus image classification	SVM	~	96% area under the ROC curve
Voets et al. (2019)	Fundus image classification	CNN	~	95.1% area under the ROC curve
Singh and Gorantla (2020)	Diabetic macular edema diagnosis	DMENet	~	96.6% Accuracy
Dagliati et al. (2018)	Diabetic complications prediction	SVM, RF	Logistic regression	83.8% Accuracy
Hathaway et al (2019)	DM detection	Classification and Regression Trees	Linear regression, LDA, KNN, SVM, NB	71.5% area under the ROC curve
Cao et al. (2018)	DR detection	SVM	RF, NN	98.5% area under the ROC curve
Ling et al. (2016)	Hyperglycemia detection	Extreme learning NN	Particle swarm optimization-NN, multiple regression fuzzy inference system, fuzzy inference system, Linear multiple regression	78% Sensitivity, 60% Specificity.
Tsao et al. (2018)	DR detection	SVM	Linear regression, ANN, DT	83.9% area under the ROC curve
Abbas et al. (2019)	Type-II DM prediction	SVM-RBF	Linear SVM	96.8% Accuracy
Nguyen et al. (2019)	Type-II DM prediction	Ensemble model	~	84.2% area under the ROC curve
Lai et al. (2019)	DM prediction	Gradient boosting machine	LR, RF	84.7% area under the ROC curve

The performance metric that is considered to be widely employed by researchers is Accuracy (17 times) followed by the combination of Accuracy, Sensitivity, and Specificity. This combination is utilized 19 times out of a total of 107 studies. The combination of Accuracy, Specificity, Sensitivity, and AUC is used in 9 articles. Some studies utilized Recall in place of Sensitivity, in this analysis Recall is included in Sensitivity instead of using it as a separate performance indicator.

Performance matrices that have commonly been utilized by researchers include Accuracy (in 52 articles), Sensitivity (in 37 articles), Specificity (in 37 articles), and AUC (in 24 articles). Other performance matrices that are not commonly utilized by scientists include, F-Score (in 10 articles), Precision (in 8 articles), Kappa Score (in 3 articles), Positive Predictive Value (in 3 articles), and Negative Predictive Value (in 2 articles).

Table 5
Review of Intelligent DM Assistant used by the researchers.

Reference	Intelligent assistant	Technique used
Wijesinghe et al (2019)	IDA: Automated diagnosis and prediction of severity of DR and foot ulcers	CNN
Maharjan et al. (2019)	Alexa: Food and Nutrition recommendation system	Forward and Backward Chaining
Czmil et al. (2019)	Pedometer: Records the step numbers. Accelerometers: Detect the acceleration of body movement. ActiGraph: Saves all physical activity forms, like climbing or doing push-ups.	RF
Behera et al. (2019), Esfahani et al. (2018)	E-Nose: A Non-Invasive Method for Breath to analyze of DM patient.	ANN
Al-Taee et al. (2016)	Robot assistant: Management of DM in children.	Capillary network
Neerincx et al. (2019), Sinoo et al. (2018), Henkemans et al. (2017), Neerincx et al. (2016)	Personal Assistant for a healthy Lifestyle, PAL: Supports the everyday DM supervision procedures of children, between the ages of 7 and 14 years.	A hybrid AI approach: The combination of symbolic reasoning approaches with ML methods. A mixture of Gaussian processing, collaborative filtering, and covariance matrices are utilized to keep track of child's knowledge level, and a Gated Recurrent Unit model for feature extraction to keep track of track child's emotional state, a Cognitive Agent Architecture Framework is utilized to deliver adaptive belief, goal, emotion-based clarifications, and a dialogue managing framework for the human-agent discussions.
Setiawan and Hammid (2019)	Robot Arm: reflexology treatment for diabetics	Servo motor
Yeh et al. (2017), Signore et al. (2019), Yeh et al. (2017)	Robotic pancreas transplantation: robotic pancreas replacement in type-I DM patient with class III overweightness.	Invasive robotic-assisted approach
Cañamero and Lewis (2016)	Robin: a cognitively and motivationally autonomous affective robot toddler with "robot diabetes" to help perceived self-effectiveness, insulin management, and emotional comfort in children with DM.	Sensors: (1) sonars fitted on the robot's chest helps to avoid collision with objects, along with to sense hugs from individuals, (2) Foot bumpers to sense accidents, (3) Head touch sensor to sense hits, (4) Gyroscopes to sense when the robot has fallen, (5) Simulated sensors to sense the homeostatically-controlled essential internal variables levels. Actuators: (1) running actions related to behaviors and requirements, and (2) to provide its internal state information. 4 types of Action selection: (1) Hunger: Hunt for food items, (2) Socialize: Hunt for the person hug, (3) Fatigue: Take a seat and rest/express fatigue, and (4) Play: Explore/ dance.
Blanson Henkemans et al. (2017)	Charlie: A robot buddy for children with DM for cognitive, social, and affective help for self-management.	Robots and health apps: for group or individual activities. Authoring and control: for selecting and tailoring of robot behaviors and apps. Monitor and inform: for being aware of understanding of the child's condition and behavior.
Mall et al. (2017)	Robot assistant: Diet Monitoring and Management of Diabetic Patient.	Medical sensors for Blood Glucose monitoring, Blood pressure monitoring, Pulse rate monitoring, Weight measurement. Network sensors like the Wi-Fi module and Bluetooth. Dietary sensors like EMG Sensor. The security part is maintained by HTTP/HPPTS protocol.
Rajalakshmi et al. (2018)	Smartphone-based system: Automatic detection of DR	The machine first trained with some annotated fundus images. After training it is tested with test samples to determine which patients with DR need doctor's consultation.
Boubin and Shrestha (2019)	Microcontroller based agent	Microcontroller based agent is used to measure the blood glucose level from the patient's volatile organic compound.
Choudhary et al. (2019)	Sensor Augmented Pump (SAP) therapy: Automatic insulin pump	Statistical analysis of continuous glucose monitoring data to automate insulin pump therapy.
Steinert et al. (2017), Dió et al. (2015)	Smartphone app: Self-monitoring daily.	The system is trained using recorded heath data of patients like physical activity, blood sugar level, weight, etc. for 3 months.
Zhang et al. (2019)	Gold Nanoclusters (AuNCs) and diabetic mice: Automated insulin-releasing agent	Bovine Serum Albumin (BSA) and statistical analysis of recorded data.
Raiff et al. (2016)	Text Message reminder: Oral Medication Reminder	The system is trained with questionnaires on diabetic specific history, general health history, and the details of the patient's medication before using the device.

9. Discussion

This research provided a review of ML-based techniques and AI-based techniques used to build automatic DM detection systems. Primary articles are chosen from the Scopus and PubMed research database. Also, two separate filters are used to choose the standard primary research articles for this analysis and to reduce the prejudice of the selection study. A total of 107 research articles is chosen for this review after a thorough screening. This review addresses chosen articles focused on five specific factors, including databases, ML-related DM detection and diagnosis, classification and diagnosis methods, AI-driven intelligent assistants for DM patients, and performance metrics used to assess the efficiency of the classification model.

Researchers used either self-created, unique datasets or publicly available datasets in the selected articles (see Table 1). In a variety of articles, several researchers have established their own unique (self-created) datasets obtained from some particular system or hospital. Our analysis showed that a variety of experiments utilized unique datasets. However, there are two significant flaws in these experiments. First, the established classification model is focused on a specific modality in which the data are obtained from one hospital and extracted from one specific tool. Hence, the classification model learned from the data obtained could not be implemented on a larger scale. The reason behind this is there are so many diagnostic tools that collect DM data these days. Each system can have its standard with various features and parameters. It is also recommended that data on multimodal DM should

Table 6
Performance measures utilized in the selected research articles.

Acc	SEN	SPE	AUC	FSc	Prec	NPV	KSc	PPV	References
✓	✗	✗	✗	✗	✗	✗	✗	✗	Mohebbi et al., 2017; Ghani et al., 2019; Jahangir et al., 2017; Huang et al., 2018; Sah and Sarma, 2018; Chen et al., 2019; He et al., 2019; Qomariah et al., 2019; Arora and Pandey, 2019; Anggraeni and Wibawa, 2019; Ali et al., 2020; Swapna et al., 2018; Al-Zebari and Sengur, 2019; Shanthi et al., 2019; Kriještorac et al., 2019; Chandakkar et al., 2017; Bhuvaneswari and Manikandan, 2018.
✓	✓	✓	✗	✗	✗	✗	✗	✗	Chen et al., 2017; Sun et al., 2019; Singh and Gupta, 2019; Aminah and Saputro, 2019; Samant and Agarwal, 2018; Benbelkacem, S and Atmani, 2019; Somasundaram and Alli, 2017; Hassan et al., 2019; Arcadu et al., 2019; Mule et al., 2018; Benzamin and Chakraborty, 2018; Aminah and Saputro, 2019; Prabhu et al., 2019; Karkuzhali and Manimegalai, 2019; Nijalingappa and Sandeep, 2015; Jayashree and Kumar, 2019; Hao et al., 2019; Abbas et al., 2019; Rahim et al., 2015.
✓	✓	✓	✓	✗	✗	✗	✗	✗	Xu and Wang, 2019; Sneha and Gangil, 2019; Kanungo et al., 2017; Lekha and Suchetha, 2017; Chen et al., 2015; Li et al., 2019; Jaya et al., 2015; Jiang et al., 2019; Tsao et al., 2018.
✓	✓	✓	✗	✗	✗	✗	✓	✗	Kwasigroch et al., 2018
✗	✓	✓	✓	✓	✗	✓	✗	✓	Sparapani et al., 2019
✓	✗	✓	✗	✓	✓	✗	✗	✗	Karthikeyan et al., 2019
✗	✓	✓	✗	✗	✗	✗	✗	✗	Lokuarachchi et al., 2019; Ling et al., 2016.
✓	✗	✗	✓	✗	✗	✗	✗	✗	Chetoui et al., 2018
✓	✗	✗	✓	✓	✓	✗	✗	✗	Nazir et al., 2019
✗	✗	✗	✓	✗	✗	✗	✗	✗	He et al., 2019; Wang and Liu, 2017; Aiello et al., 2018; Sultan et al., 2020; Moreno et al., 2016; Severeyn et al., 2019; Martinez-Vernon et al., 2018
✓	✓	✓	✗	✓	✓	✗	✗	✗	Gadekallu et al., 2020
✓	✗	✗	✗	✓	✓	✗	✓	✗	Nnamoko and Korkontzelos, 2020
✗	✓	✗	✓	✗	✗	✗	✗	✓	Datta et al., 2019
✓	✓	✓	✗	✓	✓	✗	✗	✗	El-Sappagh et al., 2015; Sangeethaa and Maheswari, 2018
✓	✓	✓	✓	✗	✓	✗	✗	✗	Samant and Agarwal, 2019
✓	✓	✓	✗	✓	✗	✗	✗	✗	Yin et al., 2019
✗	✓	✓	✓	✗	✗	✗	✗	✗	Dinh et al., 2019; Nguyen et al., 2019
✗	✗	✗	✓	✗	✗	✗	✗	✗	Voets et al., 2019; Lai et al., 2019
✓	✓	✓	✗	✓	✓	✗	✓	✗	Singh and Gorantla, 2020
✓	✓	✓	✓	✗	✗	✓	✗	✓	Dagliati et al., 2018
✗	✗	✗	✓	✓	✗	✗	✗	✗	Cao et al., 2018; Hathaway et al., 2019

** Acc: Accuracy, SEN: Sensitivity, SPE: Specificity, AUC: Area Under Curve, FSc: F-Score, Prec: Precision, NPV: Negative Predictive Value, KSc: Kappa Score, PPV: Positive Predictive Value.

be collected to establish a classification architecture where data on DM should be collected from various clinics and data collected from a range of diagnostic instruments. Such a multimodal dataset can generate a more reliable classification model that can be applied on a broader scale. Second, in a variety of tests, unique databases include a limited amount of DM results. The stated classification model that therefore suffers from over-fitting or under-fitting. As a consequence, several researchers have utilized publicly available reference databases to address these weaknesses. However, in these researches, there is still a need for more precise DM detection and diagnosis systems that can boost recognition efficiency in terms of precision and time. It is therefore strongly recommended that a structured publicly accessible diabetic dataset be utilized to treat the DM disease at its stage.

Researchers have used either ML-based methods or AI-based strategies to identify and treat DM and to assist patients with DM. Of the 107 studies chosen, 79 studies have used ML-based strategies and the remaining 28 used AI-based approaches. The most critical challenge in ML-based methods is to find valuable features that can be given as inputs to ML algorithms to generate a classification architecture. Many researchers have therefore used several features (i.e., shape features, texture features, color features, and text features) to test the efficiency of the classification to determine which features may result in improved classification results. Several ML-based researches have confirmed that text, texture, and combination of shape, texture, and color features are results-oriented and discriminatory in the diagnosis of DM. Therefore, it can be inferred that text-based, texture-based, and fusion of shape-based, color-based, and texture-based features are the most biased features of DM recognition algorithms and can yield promising outcomes for scientists operating in the area of ML-based DM detection and diagnosis and classification. The probable explanations for the significance of these features may be that, in shape features, the outline and structure of different images can be extracted of diabetic retinopathy, such as soft and hard exudates, microaneurysms and hemorrhages. However, in these features, the region, and diameter, circularity, axis length, compactness, etc. can be measured of each specific lesion relevant to the retinopathy image. Also, texture and color features can be used widely as such features offer valuable details on the appearance of diabetic retinopathy images that yield successful outcomes. Also, text-based features give the information based on age, glucose level, heredity, ECG based features like slope, amplitude, and width, etc. which are very useful to identify DM at an initial stage.

In ML methods, researchers have utilized either supervised, unsupervised or reinforcement ML algorithms to identify DM. However, a variety of experiments have used supervised machine learning methods as opposed to unsupervised approaches. Since supervised ML algorithms are more reliable than unsupervised algorithms (Guerra et al., 2011). In comparison, supervised ML algorithms learn the classification rules from recognized labeled samples (i.e., training set). This can be shown from the collected papers that the supervised ML methods have produced improved classification outcomes relative to unsupervised methods. In several studies that used supervised ML methods, Deep Neural Network and Support Vector Machine reported better classification outcomes followed by Random Forest and Ensemble Classifier. The most prominent and widely utilized DL architecture in the selected 79 studies is CNN and Deep CNN, which has been utilized in several studies. It can be concluded from this that CNN is the most appropriate deep neural network especially for diabetic detection and diagnosis and usually for the treatment of all other medical problems. For diabetic detection and diagnosis where DL is used, we found that deep architectures worked well while there are fewer levels. However, if the class number rises, the efficiency of the DL system is unsatisfactory. However, in a variety of studies, authors have utilized several ML algorithms for derived feature vectors to build classification models and evaluate which algorithm achieves better performance on the datasets utilized. In comparison, some researchers have used only one ML method to build a classification model. Yet, according to no free lunch theorem (Simonyan and Zisserman, 2014), no single ML algorithm is ideal for all image types. Thus, one can test the output of many ML algorithms and see which one will generate the finest possible outcomes.

Out of 107 studies, 28 articles have mentioned the benefit of using artificial intelligent assistant for DM patients. Digitization has enabled big data technology more important in the field of healthcare. The introduction of smart apps has facilitated the transition to digital healthcare. Artificial intelligent devices also made self-management and personalization of DM accessible for patients. Sensors are capable of sending data to mobile devices and diabetics may track their blood glucose levels through such systems. AI algorithms have enabled methods for forecasting glucose that help to combat DM effectively. They will emulate the cognitive functions of humans. AI is exploiting the explosion of medical information. AI is improving health quality for people with DM. In various studies, researchers have mentioned the use of different intelligent assistants to help patients in self-management and personalization of DM especially for type-I diabetes like self-monitoring of blood glucose, blood pressure, pulse rate, etc. Some researchers have mentioned the benefit of using chatbots where the consumer asks the chatbot to count his calories and also to obtain knowledge on health and nutrition awareness from the system.

Most of the scientists utilized specificity, sensitivity, and accuracy to test the efficiency of their classifiers. This mixture is often utilized for DM prediction utilizing machine learning techniques. Another often utilized fusion by the researchers' community is Accuracy, Specificity, Sensitivity, and AUC. This combination of measures is sufficient for DL methods where there are unnecessary groups in DM evidence and the authors have balanced groups by that or choosing approximately identical numbers in diabetic data in specific levels of training and research sets.

10. Future scopes and challenges

This segment presents several scientific problems that researchers have not been able to tackle in prior diabetic detection studies. Significant work is however also required to enhance the efficiency of various diabetic detection techniques. The research challenges that need to be tackled are laid out below.

10.1. No automated optimization technique

Deep learning has typically obtained promising outcomes in the field of DM detection and diagnosis, but the context of DL models is not fully known and is perceived to be a black box. For instance, several scientists have modified the established DL algorithms, like Deep NN or CNN, to enhance classification efficiency (Suzuki, 2017; Fu et al., 2020; Esteva et al., 2019). However, in many instances, the background of DL models is not well known and is perceived to be a black box. It is also challenging to discover the correct configuration and optimal values for the layer numbers and node numbers in various layers. Basic domain information is also necessary for the choice of values for the number of epochs, learning rate, and the strength of the regularizer. Thus, in the future, automatic optimization approaches can be introduced to determine optimum values for various DL architecture elements for specific DM datasets and other clinical datasets.

10.2. Training with inadequate data

DL software typically needs a significant number of diabetic data for training. When the training range is limited, it cannot yield sufficient results in terms of precision (Ha et al., 2018; Al-Stouhi and Reddy, 2016). There are two approaches to this problem. First, using low learning algorithms to capture training data. Second, using a range of enhancing approaches, including cropping, rotating, flipping, and color casting. Further investigations are required to produce more detailed training data, such that the DL design can be trained with reliability and more distinguishing features.

10.3. The integration of DL, AI, and cloud computing

In general, rural areas struggle from a shortage of human resources, particularly in medicine. Therefore, in these situations, AI may play a crucial role in resolving this constraint in the context of telemedicine. DL, AI, and cloud computing will be combined in the future to diagnose DM (Rajalakshmi et al., 2018; Chen et al., 2018). For example, in rural areas, the patient can utilize his or her smart insulin pen which can automatically record insulin doses, tracks active insulin, recommend mealtime and correction doses, share therapy data with your doctor or caregiver using a cloud computing environment. This implemented model will then identify the patient's condition from the data uploaded and send the findings of the diagnosis and referral back to the patient.

11. Conclusions

This paper provided a thorough study of automatic diabetic detection and diagnosis techniques. The main articles are collected from Scopus and PubMed scientific repositories. After a thorough screening process, 107 studies are chosen for this study. Each research is addressed in this analysis from the viewpoint of four specific aspects, including databases, ML-based classification, and diagnostic methods, AI-based intelligent assistants for patients with DM, and performance metrics. Several publicly accessible databases with specific characteristics have been described and documented in this study. Among these datasets, Pima Indians Diabetes Dataset, DIARETDB1, DIARETDB0, Kaggle, STARE, and Messidor are most commonly utilized for DM detection. Text, shape, and texture feature produced better outcomes. In ML algorithms, most of the studies stated that Deep Neural Network and Support Vector Machine delivers better classification outcomes followed by random forest and Ensemble Classifier. CNN is mainly used in deep learning to automatically retrieve and identify DM data. Many researchers have developed different intelligent assistants like chatbots and robots which can be used to supports the daily DM management processes of patients like insulin management, diet monitoring, etc. In terms of performance assessment, the majority of scientists utilized the accuracy, specificity, sensitivity, and AUC as indicators. The importance of the results of the analysis is addressed in a distinct discussion segment. This review also presented three new research challenges in the DM detection and diagnosis field. We hope that this thorough analysis should offer a deep overview of the area of DM detection methods and will also give useful guidance to researchers employed in this sector. There are two drawbacks to this study. First, only papers written between January 2015 and March 2020 have been included in this study. Second, it might be that the authors have overlooked certain valuable keywords and certain bibliographic sources that might have some relevant papers. Therefore, in the future, the reach of the study can be expanded to address such limitations. Lastly, it is hoped that this analysis may prove beneficial for researchers employed in the field of automatic DM detection, diagnosis, self-management, and personalization of DM therapy.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Sparapani, R., Dabbouseh, N.M., Gutterman, D., Zhang, J., Chen, H., Bluemke, D.A., Joao, A.C.L., Gregory, L.B., Soliman, E.Z., 2019. Detection of left ventricular hypertrophy using bayesian additive regression trees: The MESA (Multi-Ethnic Study of Atherosclerosis). J. Am. Heart Assoc. 8, (5) e009959.

Aiello, E.M., Toffanin, C., Messori, M., Cobelli, C., Magni, L., 2018. Postprandial glucose regulation via KNN meal classification in type 1 diabetes. IEEE Control Syst. Lett. 3 (2), 230–235.

Sharma, N., Singh, A., 2018. Diabetes detection and prediction using machine learning/IoT: A SURVEY. In: Springer International Conference on Advanced Informatics for Computing Research, pp. 471–479.

Natarajan, S., Jain, A., Krishnan, R., Rogye, A., Sivaprasad, S., 2019. Diagnostic accuracy of community-based diabetic retinopathy screening with an offline artificial intelligence system on a smartphone. JAMA Ophthalmol. 137 (10), 1182–1188.

Ramsingh, J., Bhuvaneswari, V., 2018. An efficient map reduce-based hybrid NBC-TFIDF algorithm to mine the public sentiment on diabetes mellitus – a big data approach. J. King Saud Univ.-Comput. Inform. Sci.in Press.

Carracedo, J., Alique, M., Ramírez-Carracedo, R., Bodega, G., Ramírez, R., 2019. Endothelial extracellular vesicles produced by senescent cells: pathophysiological role in the cardiovascular disease associated with all types of diabetes mellitus. Curr. Vasc. Pharmacol. 17 (5), 447–454.

Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., Simonyan, K., Hassabis, D., Lillicrap, T., 2018. A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. Science 362 (6419), 1140–1144.

Krotov, D., Hopfield, J.J., 2019. Unsupervised learning by competing hidden units. Proc. Natl. Acad. Sci. 116 (16), 7723–7731.

Taherkhani, A., Belatreche, A., Li, Y., Maguire, L.P., 2018. A supervised learning algorithm for learning precise timing of multiple spikes in multilayer spiking neural networks. IEEE Trans. Neural Networks Learn. Syst. 29 (11), 5394–5407.

Chatrati, S.P., Hossain, G., Goyal, A., Bhan, A., Bhattacharya, S., Gaurav, D., Tiwari, S. M., 2020. Smart home health monitoring system for predicting type 2 diabetes and hypertension. J. King Saudi Univ.-Comput. Inform. Sci. in press.

Sosale, B., Aravind, S.R., Murthy, H., Narayana, S., Sharma, U., Gowda, S.G., Naveenam, M., 2020. Simple, mobile-based artificial intelligence algorithm in the detection of diabetic retinopathy (SMART) study. BMJ Open Diabetes Res. Care 8 (1).

El-Sappagh, S., Ali, F., El-Masri, S., Kim, K., Ali, A., Kwak, K.S., 2018. Mobile health technologies for diabetes mellitus: current state and future challenges. IEEE Access 7, 21917–21947.

Thompson, D., Baranowski, T., 2019. Chatbots as extenders of pediatric obesity intervention: an invited commentary on "feasibility of pediatric obesity & prediabetes treatment support through tess, the ai behavioral coaching chatboT". Transl. Behav. Med. 9 (3), 448–450.

Al-Taee, M.A., Kapoor, R., Garrett, C., Choudhary, P., 2016a. Acceptability of robot assistant in management of type 1 diabetes in children. Diabetes Technol. Ther. 18 (9), 551–554.

Srividhya, E., Muthukumaravel, A., 2019. Diagnosis of diabetes by tongue analysis. In: 2019 1st IEEE International Conference on Advances in Information Technology (ICAIT), pp. 217–222.

Manikandan, K., 2019. Diagnosis of diabetes diseases using optimized fuzzy rule set by grey wolf optimization. Pattern Recogn. Lett. 125, 432–438.

Qureshi, I., Ma, J., Abbas, Q., 2019. Recent development on detection methods for the diagnosis of diabetic retinopathy. Symmetry 11 (6), 749.

Sun, Y.L., Zhang, D.L., 2019. Machine learning techniques for screening and diagnosis of diabetes: a survey. Tehnički vjesnik. 26 (3), 872–880.

Choudhury, A., Gupta, D., 2019. A survey on medical diagnosis of diabetes using machine learning techniques. In: Springer Recent Developments in Machine Learning and Data Analytics, pp. 67–78.

Raman, V., Then, P., Sumari, P., 2016. Proposed retinal abnormality detection and classification approach: Computer aided detection for diabetic retinopathy by machine learning approaches. In: 2016 8th IEEE International Conference on Communication Software and Networks (ICCSN), pp. 636–641.

Artzi, N.S., Shilo, S., Hadar, E., Rossman, H., Barbash-Hazan, S., Ben-Haroush, A., Balicer, R.D., Feldman, B., Wiznitzer, A., Segal, E., 2020. Prediction of gestational diabetes based on nationwide electronic health records. Nat. Med. 26 (1), 71–76.

Cruz-Vega, I., Peregrina-Barreto, H., de Jesus Rangel-Magdaleno, J., Ramirez-Cortes, J.M., 2019. A comparison of intelligent classifiers of thermal patterns in diabetic foot. In: 2019 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), pp. 1–6.

Theera-Umpon, N., Poonkasem, I., Auephanwiriyakul, S., Patikulsila, D., 2019. Hard exudate detection in retinal fundus images using supervised learning. Neural Comput. Appl., 1–18

Afzali, S., Yildiz, O., 2018. An effective sample preparation method for diabetes prediction. Int. Arab J. Inf. Technol. 15 (6), 968–973.

Gupta, A., Chhikara, R., 2018. Diabetic retinopathy: Present and past. Proc. Comput. Sci. 132, 1432–1440.

Chandakkar, P.S., Venkatesan, R., Li, B., 2017. MIRank-KNN: multiple-instance retrieval of clinically relevant diabetic retinopathy images. J. Med. Imaging 4, (3) 034003.

Qomariah, D.U.N., Tjandrasa, H., Fatichah, C., 2019. Classification of Diabetic Retinopathy and Normal Retinal Images using CNN and SVM. In: 2019 12th

International Conference on Information & Communication Technology and System (ICTS), pp. 152–157.
Mohebbi, A., Aradóttir, T.B., Johansen, A.R., Bengtsson, H., Fraccaro, M., Mørup, M., 2017. A deep learning approach to adherence detection for type 2 diabetics. In: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 2896–2899.
Chen, W., Chen, S., Zhang, H., Wu, T., 2017. A hybrid prediction model for type 2 diabetes using K-means and decision tree. In: 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS), pp. 386–390.
Somasundaram, S.K., Alli, P., 2017. A machine learning ensemble classifier for early prediction of diabetic retinopathy. J. Med. Syst. 41 (12), 201.
Xu, Z., Wang, Z., 2019. A risk prediction model for type 2 diabetes based on weighted feature selection of random forest and XGBoost ensemble classifier. In: 2019 IEEE Eleventh International Conference on Advanced Computational Intelligence (ICACI), pp. 278–283.
Ghani, A., See, C.H., Sudhakaran, V., Ahmad, J., Abd-Alhameed, R., 2019. Accelerating retinal fundus image classification using artificial neural networks (ANNs) and reconfigurable hardware (FPGA). Electronics 8 (12), 1522.
M. Jahangir H. Afzal M. Ahmed K. Khurshid R. Nawaz Jahangir, M., Afzal, H., Ahmed, M., Khurshid, K., Nawaz, R., 2017. An expert system for diabetes prediction using auto tuned multi-layer perceptron. In: 2017 IEEE Intelligent Systems Conference (IntelliSys), pp. 722–728.
Sneha, N., Gangil, T., 2019. Analysis of diabetes mellitus for early prediction using optimal features selection. J. Big Data 6 (1), 13.
Hassan, B., Ahmed, R., Li, B., Hassan, O., Hassan, T., 2019. Automated retinal edema detection from fundus and optical coherence tomography scans. In: 2019 IEEE 5th International Conference on Control, Automation and Robotics (ICCAR), pp. 325–330.
Guerra, L., McGarry, L.M., Robles, V., Bielza, C., Larrañaga, P., Yuste, R., 2011. Comparison between supervised and unsupervised classifications of neuronal cell types: a case study. Dev Neurobiol. 71 (1), 71–82.
Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556
Kwasigroch, A., Jarzembinski, B., Grochowski, M., 2018. Deep CNN based decision support system for detection and assessing the stage of diabetic retinopathy. In: 2018 IEEE International Interdisciplinary PhD Workshop (IIPhDW), pp. 111–116.
Sah, P., Sarma, K.K., 2018. Bloodless technique to detect diabetes using soft computational tool. In: IGI Global Ophthalmology: Breakthroughs in Research and Practice, pp. 34–52.
Woldaregay, A.Z., Årsand, E., Botsis, T., Albers, D., Mamykina, L., Hartvigsen, G., 2019. Data-driven blood glucose pattern classification and anomalies detection: machine-learning applications in type 1 diabetes. J. Med. Internet Res. 21, (5) e11030.
Arora, M., Pandey, M., 2019. Deep Neural Network for Diabetic Retinopathy Detection. In: 2019 IEEE International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), pp. 189–193.
Kanungo, Y.S., Srinivasan, B., Choudhary, S., 2017. Detecting diabetic retinopathy using deep learning. In: 2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), pp. 801–804.
D.B. Mule S.S. Chowhan D.R. Somwanshi Mule, D.B., Chowhan, S.S., Somwanshi, D.R., 2018. Detection and classfication of non-proliferative diabetic retinopathy using retinal images. In: Springer International Conference on Recent Trends in Image Processing and Pattern Recognition, pp. 312–320.
Benzamin, A., Chakraborty, C., 2018. Detection of Hard Exudates in Retinal Fundus Images Using Deep Learning. In: IEEE 2018 Joint 7th International Conference on Informatics, Electronics & Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision & Pattern Recognition (icIVPR), pp. 465–469.
Karthikeyan, S., Kumar, P.S., Madhusudan, R.J., Sundaramoorthy, S.K., Namboori, P. K., 2019. Detection of multi-class retinal diseases using artificial intelligence: an expeditious learning using deep cnn with minimal data. Biomed. Pharmacol. J. 12 (3), 1577–1586.
Lokuarachchi, D., Muthumal, L., Gunarathna, K., Gamage, T.D., 2019. Detection of red lesions in retinal images using image processing and machine learning techniques. In: IEEE 2019 Moratuwa Engineering Research Conference (MERCon), pp. 550–555.
Sokolova, M., Lapalme, G., 2009. A systematic analysis of performance measures for classification tasks. Inf. Process. Manage. 45 (4), 427–437.
Chen, G., Faris, P., Hemmelgarn, B., Walker, R.L., Quan, H., 2009. Measuring agreement of administrative data with chart data using prevalence unadjusted and adjusted kappa. BMC Med. Res. Methodol. 9, 5.
Anggraeni, Z., Wibawa, H.A., 2019. Detection of the emergence of exudate on the image of retina using extreme learning machine method. In: IEEE 2019 3rd International Conference on Informatics and Computational Sciences (ICICoS), pp. 1–6.
Ali, A., Alrubei, M.A., Hassan, L.F.M., Al-Ja'afari, M.A., Abdulwahed, S.H., 2020. diabetes classification based on KNN. IIUM Eng. J. 21 (1), 175–181.
Aminah, R., Saputro, A.H., 2019. Diabetes prediction system based on iridology using machine learning. In: IEEE 2019 6th International Conference on Information Technology, Computer and Electrical Engineering (ICITACEE), pp. 1–6.
Swapna, G., Vinayakumar, R., Soman, K.P., 2018. Diabetes detection using deep learning algorithms. ICT Express. 4 (4), 243–246.
Sisodia, D.S., Nair, S., Khobragade, P., 2017. Diabetic retinal fundus images: preprocessing and feature extraction for early detection of Diabetic Retinopathy. Biomed. Pharmacol J. 10 (2), 615–626.
Chetoui, M., Akhloufi, M.A., Kardouchi, M., 2018. Diabetic retinopathy detection using machine learning and texture features. In: 2018 IEEE Canadian Conference on Electrical & Computer Engineering (CCECE), pp. 1–4.
Nazir, T., Irtaza, A., Shabbir, Z., Javed, A., Akram, U., Mahmood, M.T., 2019. Diabetic retinopathy detection through novel tetragonal local octa patterns and extreme learning machines. Artif. Intell. Med. 99, 101695.
Prabhu, N., Bhoir, D., Shanbhag, N., 2019. Diabetic retinopathy screening using machine learning for hierarchical classification. Int. J. Innov. Technol. Exploring Eng. 8 (10), 1943–1948.
Karkuzhali, S., Manimegalai, D., 2019. Distinguising proof of diabetic retinopathy detection by hybrid approaches in two dimensional retinal fundus images. J. Med. Syst. 43 (6), 173.
Wang, Y., Liu, Z.P., 2017. Identifying biomarkers of diabetes with gene coexpression networks. In: IEEE 2017 Chinese Automation Congress (CAC), pp. 5283–5286.
Nijalingappa, P., Sandeep, B., 2015. Machine learning approach for the identification of diabetes retinopathy and its stages. In: IEEE 2015 International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), pp. 653–658.
Benbelkacem, S., Atmani, B., 2019. Random forests for diabetes diagnosis. In: IEEE 2019 International Conference on Computer and Information Sciences (ICCIS), pp. 1–4.
Lakshmi, V.S., Nithya, V., Sripriya, K., Preethi, C., Logeshwari, K., 2019. Prediction of diabetes patient stage using ontology based machine learning system. In: 2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN), pp. 1–4.
Chen, Q., Sun, X., Zhang, N., Cao, Y., Liu, B., 2019. Mini lesions detection on diabetic retinopathy images via large scale CNN features. arXiv preprint arXiv:1911.08588.
Moreno, E.M., Luján, M.J.A., Rusinol, M.T., Fernández, P.J., Manrique, P.N., Trivino, C. A., Miquel, M.P., Rodríguez, M.A., Burguillos, M.J.G., 2016. Type 2 diabetes screening test by means of a pulse oximeter. IEEE Trans. Biomed. Eng. 64 (2), 341–351.
Kriještorac, M., Halilović, A., Kevric, J., 2019. The impact of predictor variables for detection of diabetes mellitus type-2 for Pima Indians. In: Springer International Symposium on Innovative and Interdisciplinary Applications of Advanced Technologies, pp. 388–405.
Al-Zebari, A., Sengur, A., 2019. Performance comparison of machine learning techniques on diabetes disease detection. In: IEEE 2019 1st International Informatics and Software Engineering Conference (UBMYK), pp. 1–4.
Dai, J., Ji, Z., Du, Y., Chen, S., 2018. In vivo noninvasive blood glucose detection using near-infrared spectrum based on the PSO-2ANN model. Technol. Health Care 26 (S1), 229–239.
Lekha, S., Suchetha, M., 2017. Real-time non-invasive detection and classification of diabetes using modified convolution neural network. IEEE J. Biomed. Health. Inf. 22 (5), 1630–1636.
Chen, H., Yang, B., Liu, D., Liu, W., Liu, Y., Zhang, X., Hu, L., 2015. Using blood indexes to predict overweight statuses: an extreme learning machine-based approach. PLoS ONE 10 (11).
He, K., Huang, S., Qian, X., 2019a. Early detection and risk assessment for chronic disease with irregular longitudinal data analysis. J. Biomed. Inform. 96, 103231.
Gadekallu, T.R., Khare, N., Bhattacharya, S., Singh, S., Reddy Maddikunta, P.K., Ra, I. H., Alazab, M., 2020. Early detection of diabetic retinopathy using PCA-firefly based deep learning model. Electronics 9 (2), 274.
Nnamoko, N., Korkontzelos, I., 2020. Efficient treatment of outliers and class imbalance for diabetes prediction. Artif. Intell. Med. 104, 101815.
Arcadu, F., Benmansour, F., Maunz, A., Michon, J., Haskova, Z., McClintock, D., Adamis, A.P., Willis, J.R., Prunotto, M., 2019. Deep learning predicts OCT measures of diabetic macular thickening from color fundus photographs. Invest. Ophthalmol. Vis. Sci. 60 (4), 852–857.
Konstantakopoulos, F., Georga, E.I., Klampanas, K., Rouvalis, D., Ioannou, N., Fotiadis, D.I., 2019. Automatic estimation of the nutritional composition of foods as part of the GlucoseML type 1 diabetes self-management system. In: 2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE), pp. 470–473.
Huang, H., Ma, H., van Triest, H.J., Wei, Y., Qian, W., 2018. Automatic detection of neovascularization in retinal images using extreme learning machine. Neurocomputing 277, 218–227.
Datta, S., Schraplau, A., da Cruz, H.F., Sachs, J.P., Mayer, F., Böttinger, E., 2019. A machine learning approach for non-invasive diagnosis of metabolic syndrome. In: 2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE), pp. 933–940.
Bhuvaneswari, G., Manikandan, G., 2018. A novel machine learning framework for diagnosing the type 2 diabetics using temporal fuzzy ant miner decision tree classifier with temporal weighted genetic algorithm. Computing 100 (8), 759–772.
El-Sappagh, S., Elmogy, M., Riad, A.M., 2015. A fuzzy-ontology-oriented case-based reasoning framework for semantic diabetes diagnosis. Artif. Intell. Med. 65 (3), 179–208.
Sangeethaa, S.N., Maheswari, P.U., 2018. An intelligent model for blood vessel segmentation in diagnosing DR using CNN. J. Med. Syst. 42 (10), 175.
Sun, Z., Yu, S., Zhang, Y., 2019. An optimal decision tree model for diabetes diagnosis. In: IEEE 2019 4th International Conference on Computational Intelligence and Applications (ICCIA), pp. 83–87.

Samant, P., Agarwal, R., 2019. Analysis of computational techniques for diabetes diagnosis using the combination of iris-based features and physiological parameters. Neural Comput. Appl. 31 (12), 8441–8453.

Singh, A., Gupta, G., 2019. ANT_FDCSM: A novel fuzzy rule miner derived from ant colony meta-heuristic for diagnosis of diabetic patients. J. Intell. Fuzzy Syst. 36 (1), 747–760.

Aminah, R., Saputro, A.H., 2019. Application of machine learning techniques for diagnosis of diabetes based on iridology. In: IEEE 2019 International Conference on Advanced Computer Science and information Systems (ICACSIS), pp. 133–138.

Carter, J.A., Long, C.S., Smith, B.P., Smith, T.L., Donati, G.L., 2019. Combining elemental analysis of toenails and machine learning techniques as a non-invasive diagnostic tool for the robust classification of type-2 diabetes. Expert Syst. Appl. 115, 245–255.

Samant, P., Agarwal, R., 2018. Machine learning techniques for medical diagnosis of diabetes using iris images. Comput. Methods Programs Biomed. 157, 121–128.

He, B., Shu, K.I., Zhang, H., 2019b. Machine learning and data mining in diabetes diagnosis and treatment. IOP Conf. Ser.: Mater. Sci. Eng. 490, (4) 042049.

Jayashree, J., Kumar, S.A., 2019. Hybrid swarm intelligent redundancy relevance (RR) with convolution trained compositional pattern neural network expert system for diagnosis of diabetes. Health Technol., 1–10

Sultan, L.R., 2020. Diagnosis of type II diabetes based on feed forward neural network techniques. Int. J. Res. Pharm. Sci. 11 (1), 1109–1116.

Severeyn, E., Wong, S., Velásquez, J., Perpiñán, G., Herrera, H., Altuve, M., Díaz, J., 2019. Diagnosis of type 2 diabetes and pre-diabetes using machine learning. In: Springer Latin American Conference on Biomedical Engineering, pp. 792–802.

Shanthi, M., Marimuthu, R., Shivapriya, S.N., Navaneethakrishnan, R., 2019. Diagnosis of diabetes using an extreme learning machine algorithm based model. In: 2019 IEEE 10th International Conference on Awareness Science and Technology (iCAST), pp. 1–5.

Yin, H., Mukadam, B., Dai, X., Jha, N., 2019. DiabDeep: pervasive diabetes diagnosis based on wearable medical sensors and efficient neural networks. IEEE Trans. Emerg. Top. Comput.

Li, Y.H., Yeh, N.N., Chen, S.J., Chung, Y.C., 2019. Computer-assisted diagnosis for diabetic retinopathy based on fundus images using deep convolutional neural network. Mobile Inf. Syst.

Wijesinghe, I., Gamage, C., Perera, I., Chitraranjan, C., 2019. A smart telemedicine system with deep learning to manage diabetic retinopathy and foot ulcers. In: IEEE 2019 Moratuwa Engineering Research Conference (MERCon), pp. 686–691.

Maharjan, B., Li, J., Kong, J., Tao, C., 2019. Alexa, What should i eat?: A personalized virtual nutrition coach for native American diabetes patients using amazon's smart speaker technology. In: 2019 IEEE International Conference on E-health Networking, Application & Services (HealthCom), pp. 1–6.

Czmil, A., Czmil, S., Mazur, D., 2019. A method to detect type 1 diabetes based on physical activity measurements using a mobile device. Appl. Sci. 9 (12), 2555.

Behera, B., Joshi, R., Anil Vishnu, G.K., Bhalerao, S., Pandya, H.J., 2019. Electronic-nose: A non-invasive technology for breath analysis of diabetes and lung cancer patients. J. Breath Res..

Al-Taee, M.A., Al-Nuaimy, W., Muhsin, Z.J., Al-Ataby, A., 2016b. Robot assistant in management of diabetes in children based on the Internet of things. IEEE Internet Things J. 4 (2), 437–445.

Neerincx, M.A., van Vught, W., Henkemans, O.B., Oleari, E., Broekens, J., Peters, R., Frank, K., Yiannis, D., Bernd, K., Bierman, B., 2019. Socio-cognitive engineering of a robotic partner for child's diabetes self-management. Front. Rob. AI. 6.

Setiawan, A.B., Hammid, A., 2019. Simulation of robot arm for diabetes mellitus patients. J. Phys. Conf. Ser. 1424, (1) 012041.

Blanson Henkemans, O.A., Van der Pal, S., Werner, I., Neerincx, M.A., Looije, R., 2017. Learning with Charlie: a robot buddy for children with diabetes. In: Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, pp. 406–406.

Yeh, C.C., Spaggiari, M., Tzvetanov, I., Oberholzer, J., 2017. Robotic pancreas transplantation in a type 1 diabetic patient with morbid obesity: a case report. Medicine 96 (6).

Cañamero, L., Lewis, M., 2016. Making new "New AI" friends: designing a social robot for diabetic children from an embodied AI perspective. Int. J. Social Rob. 8 (4), 523–537.

Sinoo, C., van Der Pal, S., Henkemans, O.A.B., Keizer, A., Bierman, B.P., Looije, R., Neerincx, M.A., 2018. Friendship with a robot: Children's perception of similarity between a robot's physical and virtual embodiment that supports diabetes self-management. Pat. Educ. Counselling 101 (7), 1248–1255.

Mall, S., Gupta, M., Chauhan, R., 2017. Diet monitoring and management of diabetic patient using robot assistant based on Internet of Things. In: IEEE 2017 International Conference on Emerging Trends in Computing and Communication Technologies (ICETCCT), pp. 1–8.

Henkemans, O.A.B., Bierman, B.P., Janssen, J., Looije, R., Neerincx, M.A., van Dooren, M.M., de Vries, J.L.E., van der Burg, G.J., Huisman, S.D., 2017. Design and evaluation of a personal robot playing a self-management education game with children with diabetes type 1. Int. J. Hum Comput Stud. 106, 63–76.

Neerincx, A., Sacchitelli, F., Kaptein, R., Van Der Pal, S., Oleari, E., Neerincx, M.A., 2016. Child's culture-related experiences with a social robot at diabetes camps. In: 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 485–486.

Bansal, N., 2015. Prediabetes diagnosis and treatment: a review. World J. Diabetes 6 (2), 296–303.

Katsarou, A., Gudbjörnsdottir, S., Rawshani, A., Dabelea, D., Bonifacio, E., Anderson, B.J., Jacobsen, L.M., Schatz, D.A., Lernmark, Å., 2017. Type 1 diabetes mellitus. Nat. Rev. Dis. Primers 3 (1), 1–17.

DeFronzo, R.A., Ferrannini, E., Groop, L., Henry, R.R., Herman, W.H., Holst, J.J., Hu, F. B., Kahn, C.R., Raz, I., Shulman, G.I., Simonson, D.C., Testa, M.A., Weiss, R., 2015. Type 2 diabetes mellitus. Nat. Rev. Dis. Primers 1 (1), 1–22.

Flannick, J., Johansson, S., Njølstad, P.R., 2016. Common and rare forms of diabetes mellitus: towards a continuum of diabetes subtypes. Nat. Rev. Endocrinol. 12 (7), 394–407.

Suzuki, K., 2017. Overview of deep learning in medical imaging. Radiol. Phys. Technol. 10 (3), 257–273.

Fu, Y., Lei, Y., Wang, T., Curran, W.J., Liu, T., Yang, X., 2020. Deep learning in medical image registration: a review. Phys. Med. Biol.

Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Dean, J., 2019. A guide to deep learning in healthcare. Nat. Med. 25 (1), 24–29.

Ha, J., Eun, J., Ahn, P., Shin, D.H., Kim, J., 2018. Learning convolutional neural network using data from other domains in case of insufficient data. In: Proceedings of the 2018 International Conference on Information Science and System. pp. 122–126.

Al-Stouhi, S., Reddy, C.K., 2016. Transfer learning for class imbalance problems with inadequate data. Knowl. Inf. Syst. 48 (1), 201–228.

Doreswamy, G.S., Santosh, K.J., 2018. Prediction accuracy comparison of predictive models using machine learning for diabetes data set. Int. J. Adv. Res. Comput. Sci. 9 (Special Issue 3), 86.

Rajalakshmi, R., Subashini, R., Anjana, R.M., Mohan, V., 2018. Automated diabetic retinopathy detection in smartphone-based fundus photography using artificial intelligence. Eye 32 (6), 1138–1144.

Chen, M., Yang, J., Zhou, J., Hao, Y., Zhang, J., Youn, C.H., 2018. 5G-smart diabetes: Toward personalized diabetes diagnosis with healthcare big data clouds. IEEE Commun. Mag. 56 (4), 16–23.

World Health Organization, 08.06.2020 (https://www.who.int/news-room/fact-sheets/detail/diabetes)

Signore, M.A., Rescio, G., De Pascali, C., Iacovacci, V., Dario, P., Leone, A., Quaranta, F., Taurino, A., Siciliano, P., Francioso, L., 2019. Fabrication and characterization of AlN-based flexible piezoelectric pressure sensor integrated into an implantable artificial pancreas. Sci. Rep. 9 (1), 1–11.

Boubin, M., Shrestha, S., 2019. Microcontroller implementation of support vector machine for detecting blood glucose levels using breath volatile organic compounds. Sensors 19 (10), 2283–2292.

Spaggiari, M., Tzvetanov, I.G., Di Bella, C., Oberholzer, J., 2018. Robotic pancreas transplantation. Gastroenterol. Clin. North Am. 47 (2), 443–448.

Esfahani, S., Wicaksono, A., Mozdiak, E., Arasaradnam, R.P., Covington, J.A., 2018. Non-invasive diagnosis of diabetes by volatile organic compounds in urine using FAIMS and Fox4000 electronic nose. Biosensors 8 (4), 121–134.

Burki, T., 2015. Robots and type 1 diabetes. Lancet Diabetes Endocrinol.. 3 (11), 844.

Choudhary, P., De Portu, S., Arrieta, A., Castañeda, J., Campbell, F.M., 2019. Use of sensor-integrated pump therapy to reduce hypoglycaemia in people with Type 1 diabetes: a real-world study in the UK. Diabetes Med. 36 (9), 1100–1108.

Steinert, A., Haesner, M., Steinhagen-Thiessen, E., 2017. App-basiertes Selbstmonitoring bei Typ-2-Diabetes. Z. Gerontol. Geriatrie 50 (6), 516–523.

Zhang, Y., Wu, M., Dai, W., Chen, M., Guo, Z., Wang, X., Tan, D., Shi, K., Xue, L., Liu, S., Lei, Y., 2019. High drug-loading gold nanoclusters for responsive glucose control in type 1 diabetes. J. Nanobiotechnol. 17 (1), 74–85.

Raiff, B.R., Jarvis, B.P., Dallery, J., 2016. Text-message reminders plus incentives increase adherence to antidiabetic medication in adults with type 2 diabetes. J. Appl. Behav. Anal. 49 (4), 947–953.

Dió, M., Deutsch, T., Biczók, T., Mészáros, J., 2015. Intelligent interpretation of home monitoring blood glucose data. Orv. Hetil. 156 (29), 1165–1173.

Dinh, A., Miertschin, S., Young, A., Mohanty, S.D., 2019. A data-driven approach to predicting diabetes and cardiovascular disease with machine learning. BMC Med. Inf. Decis. Making 19 (1), 211–226.

Hao, Y., Cheng, F., Pham, M., Rein, H., Patel, D., Fang, Y., Yan, J., Song, X., Yan, H., Wang, Y., 2019. A Noninvasive, economical, and instant-result method to diagnose and monitor type 2 diabetes using pulse wave: case-control study. JMIR mHealth uHealth 7, (4) e11959.

Martinez-Vernon, A.S., Covington, J.A., Arasaradnam, R.P., Esfahani, S., O'connell, N., Kyrou, I., Savage, R. S, 2018. An improved machine learning pipeline for urinary volatiles disease detection: diagnosing diabetes. PLoS ONE 13 (9).

Jiang, H., Yang, K., Gao, M., Zhang, D., Ma, H., Qian, W., 2019. An interpretable ensemble deep learning model for diabetic retinopathy disease classification. In: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 2045–2048.

Jaya, T., Dheeba, J., Singh, N.A., 2015. Detection of hard exudates in colour fundus images using fuzzy support vector machine-based expert system. J. Digit. Imaging 28 (6), 761–768.

Voets, M., Møllersen, K., Bongo, L.A., 2019. Reproduction study using public data of: development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. PLoS ONE 14 (6).

Singh, R.K., Gorantla, R., 2020. DMENet: diabetic macular edema diagnosis using hierarchical ensemble of CNNs. PLoS ONE 15, (2) e0220677.

Dagliati, A., Marini, S., Sacchi, L., Cogni, G., Teliti, M., Tibollo, V., De Cata, P., Chiovato, L., Bellazzi, R., 2018. Machine learning methods to predict diabetes complications. J. Diabetes Sci. Technol. 12 (2), 295–302.

Hathaway, Q.A., Roth, S.M., Pinti, M.V., Sprando, D.C., Kunovac, A., Durr, A.J., Cook, C. C., Fink, G.K., Cheuvront, T.B., Grossman, J.H., Aljahli, G.A., Taylor, A.D., Giromini, A.P., Allen, J.L., Hollander, J.M., 2019. Machine-learning to stratify diabetic patients using novel cardiac biomarkers and integrative genomics. Cardiovasc. Diabetol. 18 (1), 78–94.

Cao, W., Czarnek, N., Shan, J., Li, L., 2018. Microaneurysm detection using principal component analysis and machine learning methods. IEEE Trans. Nanobiosci. 17 (3), 191–198.

Ling, S.H., San, P.P., Nguyen, H.T., 2016. Non-invasive hypoglycemia monitoring system using extreme learning machine for Type 1 diabetes. ISA Trans. 64, 440–446.

Tsao, H.Y., Chan, P.Y., Su, E.C.Y., 2018. Predicting diabetic retinopathy and identifying interpretable biomedical features using machine learning algorithms. BMC Bioinf. 19 (9), 283.

Abbas, H.T., Alic, L., Erraguntla, M., Ji, J.X., Abdul-Ghani, M., Abbasi, Q.H., Qaraqe, M. K., 2019. Predicting long-term type 2 diabetes with support vector machine using oral glucose tolerance test. PLoS ONE 14 (12).

Nguyen, B.P., Pham, H.N., Tran, H., Nghiem, N., Nguyen, Q.H., Do, T.T., Tran, C.T., Simpson, C.R., 2019. Predicting the onset of type 2 diabetes using wide and deep learning with electronic health records. Comput. Methods Programs Biomed. 182, 105055.

Lai, H., Huang, H., Keshavjee, K., Guergachi, A., Gao, X., 2019. Predictive models for diabetes mellitus using machine learning techniques. BMC Endocrine Disorders 19 (1), 1–9.

Donsa, K., Spat, S., Beck, P., Pieber, T.R., Holzinger, A., 2015. Towards personalization of diabetes therapy using computerized decision support and machine learning: some open problems and challenges. Smart Health, 237–260.

Vaskovsky, A.M., Chvanova, M.S., 2019. Designing the neural network for personalization of food products for persons with genetic president of diabetic sugar. In: 2019 3rd School on Dynamics of Complex Networks and their Application in Intellectual Robotics (DCNAIR), pp. 175–177.

Demongeot, J., Elena, A., Taramasco, C., Vuillerme, N., 2015. Serious games and personalization of the therapeutic education. In: International Conference on Smart Homes and Health Telematics, pp. 270–281.

Rahim, S.S., Palade, V., Jayne, C., Holzinger, A., Shuttleworth, J., 2015. Detection of diabetic retinopathy and maculopathy in eye fundus images using fuzzy image processing. In: International Conference on Brain Informatics and Health, pp. 379–388.

Zou, Q., Qu, K., Luo, Y., Yin, D., Ju, Y., Tang, H., 2018. Predicting diabetes mellitus with machine learning techniques. Front. Genet. 9, 515.

Alghamdi, M., Al-Mallah, M., Keteyian, S., Brawner, C., Ehrman, J., Sakr, S., 2017. Predicting diabetes mellitus using SMOTE and ensemble machine learning approach: The Henry Ford ExercIse Testing (FIT) project. PLoS ONE 12 (7).

Xiong, Z., Liu, T., Tse, G., Gong, M., Gladding, P.A., Smaill, B.H., Stiles, M.K., Gillis, A. M., Zhao, J., 2018. A machine learning aided systematic review and meta-analysis of the relative risk of atrial fibrillation in patients with diabetes mellitus. Front. Physiol. 9, 835.

Holzinger, A., Langs, G., Denk, H., Zatloukal, K., Müller, H., 2019. Causability and explainability of artificial intelligence in medicine. Wiley Interdiscip. Rev.: Data Min. Knowl. Disc. 9, (4) e1312.