

Leveraging Machine Learning and Computer Vision for Human Disease Detection

Mohammed Aledhari, Y. Alicia Hong, Teresa Wu, Bhavik N. Patel, Kwangsik T. Nho, Marianne Di Pierro,

Abstract—Early detection of diseases and identification of health risks is critical for reducing morbidity and mortality. Traditionally, disease screening relies on [sometimes](#) invasive clinical tests or lengthy interviews, which [may be time consuming, expensive, and inconclusive](#). In recent years, machine learning (ML) and computer vision (CV) for disease detection and risk identification provided scientific breakthroughs leading to more effective and efficient disease detection and risk identification methods (DDRI). This paper presents a comprehensive overview and roadmap for physicians, clinicians, and healthcare workers by identifying the key considerations, potential challenges and their solutions, evaluation metrics, and optimization methods for any disease detection/classification model using ML and CV approaches. In total, we cover different kinds of common diseases, where ML can contribute, including infectious and chronic conditions and mental, developmental, degenerative, neurologic disorders, as examples to illustrate the applications. In addition to common methods used in ML applications, we will also discuss enabling technologies of software and hardware, architectures, data preprocessing, optimization techniques, and evaluation metrics. We synthesize and compare previous work in ML applications, identify the gaps in the literature, and discuss future research directions.

Index Terms—Biosensors, Wearable Devices, Disease Detection, Risk Identification, Machine Learning

1 INTRODUCTION

DETERMINING effective utilization of Machine Learning (ML) methods for detecting various diseases and risk factors is an important concept in the field of biotechnology and is becoming increasingly in the medical field [1]. Identifying the presence of a disease has traditionally relied UPON equipment and/or techniques (SUCH AS) (OMIT like) handheld devices that can detect moisture to measure subepidermal pressure for pressure ulcer detection [2], abnormal glucose identification for diabetes detection [3], computerized tomography (CT) scan to affirm negative test results when SARS-CoV-2 virus (COVID-19) is suspected [4], and a variety of tests for early cardiovascular disease detection such as X-rays, electrocardiography, heart Magnetic Resonance Imaging (MRI), blood tests, etc. [5]. These methods can

be costly and sometimes invasive, resulting in the need for researchers to seek out new, easy-to-use, noninvasive, and cost-effective techniques, such as leveraging machine learning and computer vision for disease detection and risk identification (DDRI) [6].

The process of disease detection refers to revealing the presence of a disease through methods such as conditioning examinations and utilizing knowledge of specific risk factors that pertain to the disease being diagnosed [7]. It is important for physicians to identify risk factors that detect the presence of a disease as early as possible. The risk identification refers to the ability to recognize the likelihood that a person may develop a disease by examining factors such as genetics, demographics, physiology, and other factors. Risk identification takes place using robust computer procedures and statistical processes like logistic regression (LR) to select independent risk factors in large databases [8].

Biosensors work with ML algorithms to detect and measure signals that can be used for DDRI [9] by providing stability and sensitivity to evaluate biological responses in real-time and turn them into electrical signals [10]. A general representation of a biosensor is shown in Figure 1.

The interest in wearable biosensors for DDRI is increasing because these biosensors provide a non-invasive, low-cost method for obtaining patient statistics pertaining to biological conditions such as heart failure, diabetes, etc. [11] Figure 2 illustrate an example of a non-invasive

M. Aledhari is with the College of Computing and Software Engineering, Kennesaw State University, Marietta, GA, 30060 USA. Email: maledhar@kennesaw.edu

Y. Hong is with the College of Health Human Services, George Mason University, Fairfax, VA, 22030 USA. Email: yhong22@gmu.edu

T. Wu is with the School of Computing, Informatics, Decision Systems Engineering, Arizona State University, Tempe, AZ 85281 USA. Email: Teresa.Wu@asu.edu

B. Patel is with the ASU-Mayo Center for Innovative Imaging, Tempe, AZ 85281 USA. Email: Patel.Bhavik@mayo.edu

K. Nho is with the School of Informatics and Computing, Indiana University School of Medicine, Indianapolis, IN 46202 USA. Email: knho@iupui.edu

M. Di Pierro is with the Journal for Quality Perspectives in Knowledge Acquisition, Kalamazoo, MI 49008 USA. Email: marianne.dipierro@wmich.edu

Manuscript received xx, 2021; revised xx, 2021.

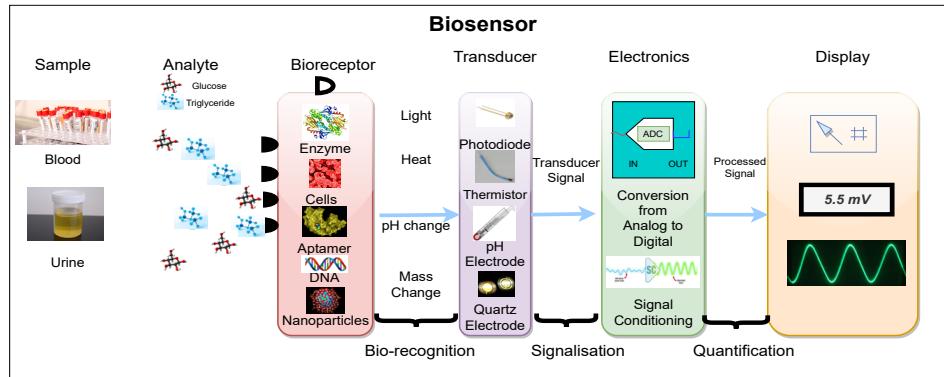


Fig. 1: Generic framework for a biosensor [9].

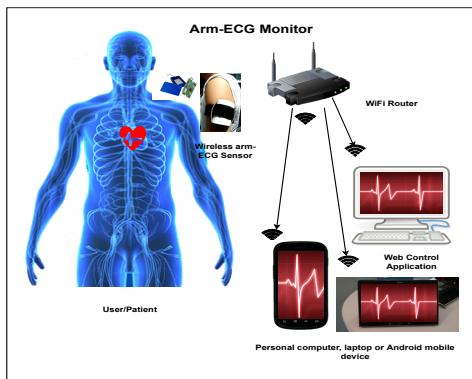


Fig. 2: Example of an arm-wearable band for heart arrhythmia detection that stores data in real-time for later download [12].

wearable device than can detect heart arrhythmias [12].

Detecting the presence of a disease and identifying risks that could potentially cause or result in a disease are important topics for research because improving the detection accuracy MAY (OMIT can) increase patient life expectancy [1]. When overall patient life expectancy increases, there is also an increase in health-related problems due to aging [13], factors which contribute to the issue of biosensor performance and desirability more important than ever. Performance improvement can be achieved by understanding and enhancing ML algorithms.

ML algorithms have been proven to identify diseases and risks more effectively than some healthcare practitioners and perform at an equal standard to healthcare experts [14]. ML algorithms can also be constructed to detect a patient's location in the event of an emergency which could potentially save the patient's life [15]. Moreover, the inclusion of ML algorithms can alleviate long wait times by automating the analysis and detection processes, leading to faster results, earlier intervention, and increased

patient satisfaction. However, several challenges are associated with the successful deployment of ML in healthcare applications.

This paper discusses the potential and the challenges associated with the deployment of ML algorithms in the early diagnose of various diseases. In detail, the paper focuses on the following categories: physical, mental, infectious, non-infectious, deficiency, inherited, degenerative, social, and self-inflicted.

1.1 Scope of the Survey

This paper aims to analyze the potential and the challenges associated with the deployment of ML algorithms in the early diagnose of various diseases. In detail, the paper focuses on nine different categories of diseases including: physical, mental, infectious, non-infectious, deficiency, inherited, degenerative, social, and self-inflicted. For physical diseases, epilepsy and chronic kidney disease are discussed in more detail. Autism and anxiety are covered from the mental diseases category. COVID-19 and Methicillin-resistant *Staphylococcus aureus* (MRSA) are included from the infectious diseases. Brain Cancer and Gastric Cancer are discussed in the non-infectious category. Phenylketonuria (PKU) and Medium-chain acyl-CoA dehydrogenase deficiency (MCADD) are deficiency diseases that are discussed, Heart Disease and Diabetes are inherited diseases that are detailed, Parkinson's Disease and Alzheimer's are discussed for degenerative diseases, HIV and HPV for social disease, and Obesity and Lung Cancer for the self-inflicted disease category.

Moreover, the paper also provides an overview of different benchmark datasets made publicly available for the research purposes. We also highlight the key research challenges and potential directions of future research.

1.2 Related Surveys

TO BE EDITED There is extensive research pertaining to the detection, diagnosis, classification, and/or risk prediction of different diseases and conditions using different

types of ML algorithms, such as statistical methods, DL, genetic algorithms (GAs), and rule-based systems. There are also some interesting surveys on the topic. Majority of the surveys focus on a particular category of diseases. For instance, Myszczyńska et al. [16] provides a detailed overview of different ML applications and their ability to detect and treat neurodegenerative diseases at early stages. Similarly, Bracher et al. [17] surveys the ML methods deployed for the prediction of psychiatric disorders (i.e., schizophrenia, bipolar, autism, and anorexia) from genetic data. An review of around 256 articles relying on ML algorithms for the detection of coronary artery disease (CAD) is provided in [18]. In [19], [20], an overview of various ML algorithms proposed in the literature for heart disease and breast cancer detection is provided. Chaudhury et al. [21] and Shatte et al. [22] provide an overview of ML algorithms proposed for diabetes and mental disorder diseases, respectively.

There are also some surveys focusing on DL based solutions only. For instance, Kieu et al. [23] discuss DL methods in the detection and categorization of medical images to detect lung diseases. They proposed a taxonomy of seven commonly observed characteristics from previous articles. Similarly, Caballe et al. [24] provides an overview of DL methods proposed for the diagnosis of various diseases.

to be conti...

1.3 Contribution

This paper provides a detailed overview of the literature aiming at the deployment of ML algorithms for disease detection. The paper mainly focuses on the use of ML algorithms in a diversified list of diseases with a particular focus on the applications, challenges, and future research directions. The paper also provides an overview of different benchmark datasets available for the training and evaluation of ML algorithms for disease detection.

The key contributions of the paper can be summarized as follows.

- We provide a detailed overview of the literature on ML in disease detection. In detail, we covered a diversified list of diseases from nine different categories.
- We also provide an overview of different benchmark datasets publicly available for the training and evaluation ML algorithms in disease detection.
- We also identify and explore the limitations, challenges, and future research direction in the domain.

The rest of the paper is organized as follows. Section yyy discusses to be added once the structure is finalized. Figure 3 visually depicts the structure of the paper.

2 DISEASES

In this section, we describe several applications of ML technologies in DDRI within the medical field for common

diseases from the following categories: physical, mental, infectious, non-infectious, deficiency, inherited, degenerative, social, and self-inflicted. For each application, the following topics will be discussed:

- The sensors used, if any
- The ML techniques employed

2.1 Epilepsy

Epilepsy is a neurological disorder that causes seizures, loss of awareness, unusual behavior, and/or sensations. Using recent clinical neurophysiology discoveries and RF-based subset feature selection from linear EEG sensor data, a RF design was constructed to detect epileptics when compared to healthy individual [40]. 99 fine-grained frequency inter-recital sub-bands were constructed from the data, the data was visually analyzed by researchers, linear decomposition was used for data cleaning, and independent component analysis was used for data preprocessing. The preprocessing technique required a linear mixed medium as well as negligible propagation delays, independent timing of the sources, and the amount of sources equal the amount of sensors used. Signals without a normal distribution are independent components, and signals with a normal distribution are heterogeneous.

For feature selection, spectral analysis was performed by gathering power per frequency [40]. The RF gives an importance score (valid above 45) to different features considering the MSE. The feature selection method was used to train a RF model for classification purposes. Out of 50 cases, only one was misclassified: one true epileptic classified as healthy. The balanced accuracy achieved was 99% with 100% sensitivity.

2.2 Chronic Kidney Disease

Chronic Kidney Disease is the gradual loss of kidney functions that causes the inability of the kidneys to properly filter blood due to waste build up. Chronic Kidney disease was diagnosed using data from the University of California's Machine Learning Repository which is considered complicated clinical data due to situations such as missing values [44]. To resolve the issue of missing values, KNN imputations was used to produce five datasets because in this type of healthcare data, certain categories depend on others (i.e., physical data for diseased persons is similar to the physical data of similarly diseased persons, and healthy persons will have similar physiological measurements). LR and RF classifiers were used to diagnose Chronic Kidney Disease with 99.83% accuracy.

For feature extraction, optimal subset regression to choose the best predictor variable combinations and RF for classification were used to select relevant variables and discard irrelevant variables [44]. Random imputation as well as mean/mode imputation were used to generate datasets to compare with the datasets using KNN imputation to show its effectiveness.

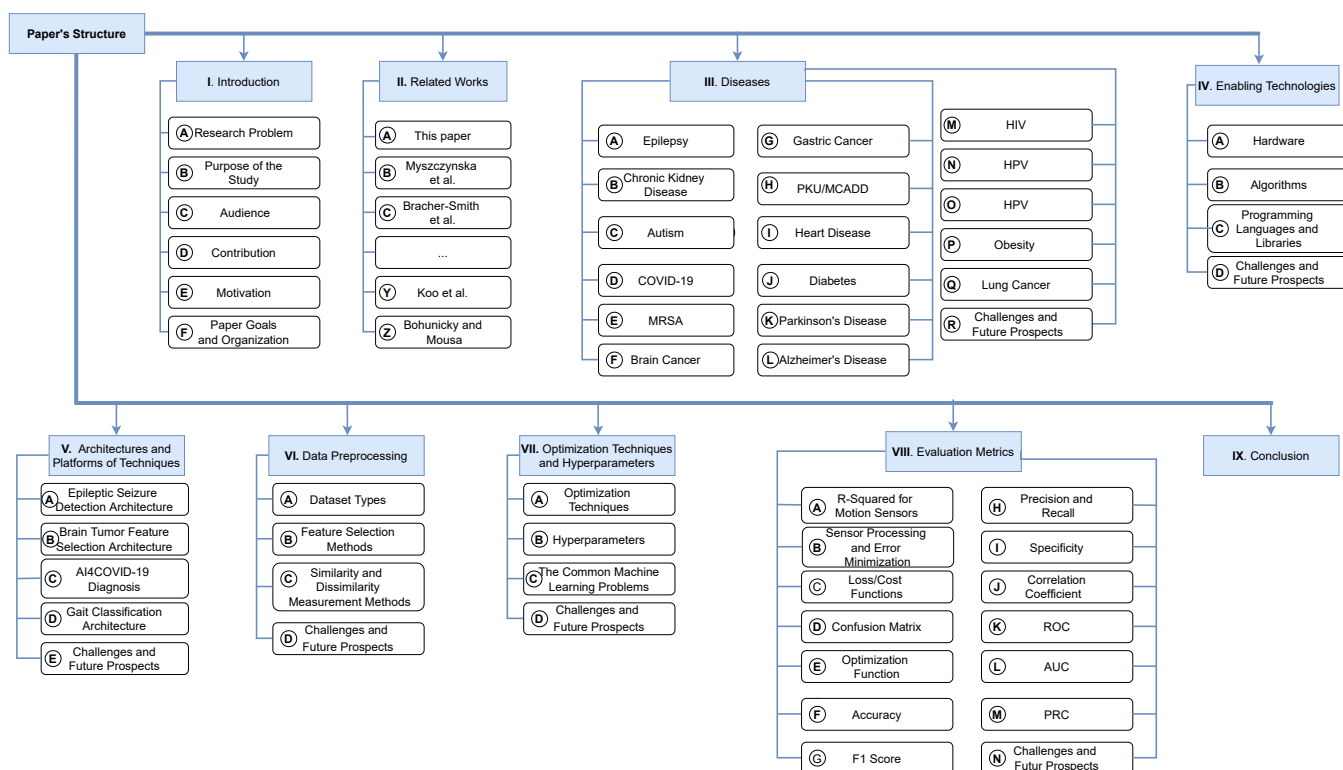
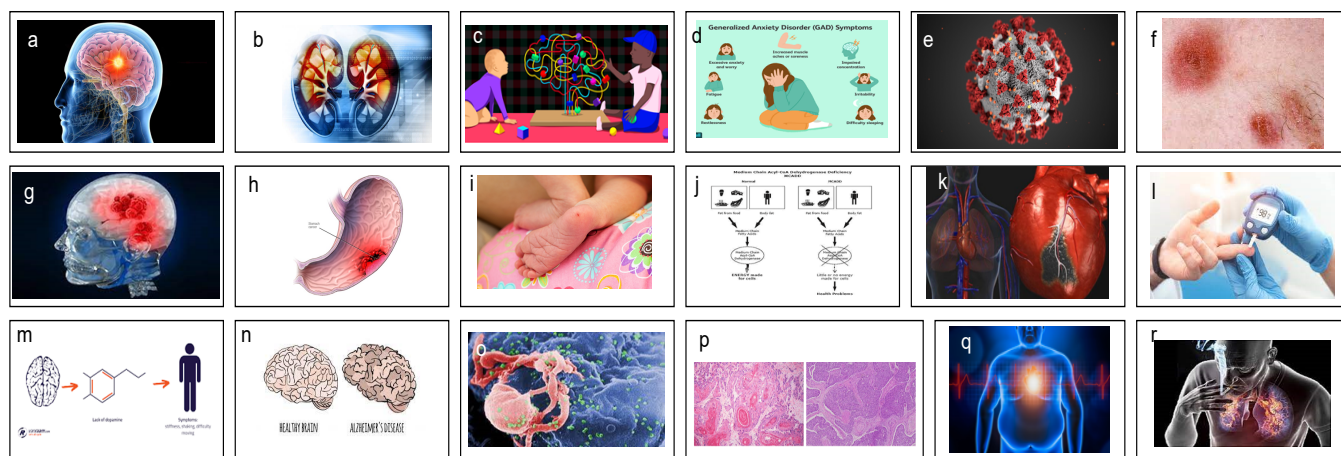
Fig. 3: Roadmap outlining the paper's organization. **TO BE FINALIZED**Fig. 4: The diseases discussed in detail in this paper: **a.** Epilepsy, **b.** Chronic Kidney Disease, **c.** Autism, **d.** Anxiety, **e.** COVID-19, **f.** MRSA, **g.** Brain Cancer, **h.** Gastric Cancer, **i.** PKU, **j.** MCADD, **k.** Heart Disease, **l.** Diabetes, **m.** Parkinson's Disease, **n.** Alzheimer's Disease, **o.** HIV, **p.** HPV, **q.** Obesity, **r.** Lung Cancer

TABLE 1: Summary of existing surveys and reviews related to ML for Disease Detection and Risk Identification

Refs. (Author)	Year of Publication	Multiple Diseases	Disease Detection	Risk Identification	Deep Learning	Wearable Devices	Evaluation	Standard- ization	Comments
This paper	2021	✓	✓	✓	✓	✓	✓	✓	ML algorithms can be used for DDRI by assisting medical professionals in accurately detecting various diseases and risks.
Myszczyńska et al. [16]	2020	✓	✓	✗	✓	✓	✓	✗	Multiple ML techniques on Neurodegenerative Diseases
Bracher-Smith et al. [17]	2020	✓	✗	✓	✓	✗	✗	✗	ML methods on the prediction of psychiatric disorders
Panicker et al. [25]	2020	✗	✓	✗	✓	✓	✓	✗	Classification of cardiovascular disease using Ensemble techniques, SVM, and CNN
Alizadehsani et al. [18]	2020	✗	✓	✗	✓	✗	✗	✗	ML and DL on CAD detection
Kieu et al. [23]	2020	✗	✓	✗	✓	✗	✗	✗	Lung disease detection using DL methods
Caballe-Cervigon et al. [24]	2020	✓	✓	✗	✓	✗	✓	✗	DL methods are most successful for handling Big Data for disease detection
Ebrahimighahnavieh et al. [26]	2020	✗	✓	✗	✓	✗	✓	✓	DL methods can be used to detect Alzheimer's Disease, but there is limited data available and training procedures can be difficult.
Saranya and Pravin [19]	2020	✓	✗	✓	✗	✓	✗	✗	Various ML models are useful at predicting Heart Disease and Breast Cancer risk
Channa et al. [27]	2020	✓	✓	✗	✗	✓	✓	✓	Wearable insoles may be best for detecting abnormal gait activity and the TRIS algorithm may be more suitable than ML algorithms.
Kahn et al. [20]	2019	✗	✓	✗	✓	✗	✗	✗	SVM, ANN, and ensemble methods are most commonly used for classifying heart diseases.
Hyde et al. [28]	2019	✗	✓	✗	✓	✗	✓	✓	Supervised ML techniques and Autism Diagnosis
Uddin et al. [29]	2019	✓	✗	✓	✗	✗	✓	✗	Discussed various Supervised ML techniques for risk identification of different diseases
Choudhury and Gupta [21]	2019	✗	✓	✗	✗	✗	✗	✗	Diabetes diagnosis using different ML classifiers
Shatte et al. [22]	2019	✓	✓	✗	✗	✗	✓	✗	ML techniques are successful at diagnosing mental health disorders
Jo et al. [30]	2019	✗	✓	✗	✓	✗	✗	✗	Traditional ML combined with SAE is effective at detecting early-onset Alzheimer's
Safdar et al. [31]	2018	✗	✓	✗	✓	✗	✓	✗	Looked at ML techniques for the classification of heart disease (NN were most commonly used).
Xiong et al. [32]	2018	✗	✗	✓	✗	✗	✓	✗	Looked at ML techniques and determined AF is a risk factor for DM patients
Pellegrini et al. [33]	2018	✗	✓	✗	✓	✗	✓	✓	Diverse data and clinically integrated ML techniques could help with future research in dementia detection.
Bakator and Radosav [34]	2018	✓	✓	✗	✓	✗	✗	✗	DL methods are superior to traditional ML methods for disease detection by image analysis
Cuadrado-Godia et al. [35]	2018	✗	✓	✗	✓	✗	✓	✗	NN and DL can be used to detect cVSD through image processing.
Shailaja et al. [36]	2018	✓	✓	✓	✗	✗	✗	✗	SVMs can effectively predict heart disease, breast cancer, and diabetes with 96.40% accuracy.

2.3 Autism

Autism Spectrum Disorder is a developmental disorder that makes social interactions and communication difficult. Autism Spectrum Disorder was classified from data obtained from the Quantitative Checklist for Autism in Toddlers (Q-CHAT) that contains a diverse sample of normally distributed autistic characteristics [63]. Supervised binary classification was used (Autism vs. Typically Developing children). The applied supervised models include: RF, NB, SVM, LR, and KNN with the highest accuracy of 95% achieved by the SVM.

The survey included 25 questions (features) about autistic behaviors. Missing values of more than 25% were eliminated during data preprocessing. Parameter tuning was then performed on the SVM, and training was then carried out. The training included fivefold cross validation, and hyperparameters were automatically tuned.

2.4 Anxiety

Anxiety is a normal stress response, but when it becomes extreme, obsessive, and disrupts daily living, it can indi-

cate an anxiety disorder. In an involving text-based data, three different mental disorders of five different severities (classes) were analyzed to predict the occurrence of depression, stress, and anxiety using ML techniques [61]. Mental disorders are more difficult to classify than other diseases using ML such as diabetes, because mental disorders have varying severity levels.

Through online questionnaires, 39,776 instances were gathered from 42 questions where the answers ranged from 1 to 4 (Did not apply at all, to Applied to me very much or most of the time). The totals for each survey were added (0 to 3 rather than 1 to 4) and the following labels were given:

- Extremely Severe anxiety: 20 or more
- Extremely Severe depression: 28 or more
- Extremely Severe stress: 33 or more
- ...
- Normal anxiety: Less than 8
- Normal depression: Less than 13
- Normal stress: Less than 15

Eight algorithms were used for classification on two

TABLE 2: Table of common physical diseases and techniques for detection.

Disease	Reference Number	Technique	ML Types	Measurement	Challenges (If available)
Sepsis	Ref. [37]	LiSep LSTM	Supervised/Unsupervised	AUR-ROC .93	1. Input features may have decreased performance
	Ref. [38]	SVM	Supervised	Precision 83.6%	2. Limited training data 1. Quick diagnosis from onset is necessary for model performance 2. Complexity
COPD	Ref. [39]	RF	Supervised	Precision 97.7%	1. Small sample size 2. Interpretability
Epilepsy	Ref. [40]	RF	Supervised	Accuracy 99%	1. Limited external validity
	Ref. [41]	MLP	Supervised	Accuracy 99.92%	2. Interpretability
	Ref. [42]	SVM	Supervised	Accuracy 94%	1. Small sample size 2. Complex 1. Small sample size 2. All parameters not optimized
Chronic Kidney Disease	Ref. [43]	DT & RF	Supervised	Accuracy 95%	1. Small sample size
	Ref. [44]	RF & LR	Supervised	Accuracy 99.83%	2. Difficult interpretation
	Ref. [45]	ANN, SVM & NB	Supervised/Unsupervised	Accuracy 98%	1. Missing values 2. Difficult interpretation
	Ref. [46]	SVM	Supervised	Accuracy 98.5%	1. Limited access to information 2. 50% of data had missing values to be removed 1. Complexity 2. Slow performance
Chronic Atrophic Gastritis	Ref. [47]	CNN	Supervised/Unsupervised	Accuracy 94.2%	1. Only clear images were used. In reality, images may be blurry. 2. Large sample size
Crohn's	Ref. [48]	CNN	Supervised/Unsupervised	Accuracy 97%	1. Slow performance 2. Works well with large datasets
Ulcerative Colitis	Ref. [49]	SVM	Supervised	Precision 100% -active cases 62% -inactive cases	1. Only two datasets for training 2. Complexity
	Ref. [50]	RF & ANN	Supervised/Unsupervised	ROC-AUC .9506	1. Limited value in predictive power 2. Small validation set
Blood Disease	Ref. [51]	Logit-Boost (Regression-based)	Supervised	Accuracy 98.16%	1.f-measures and recall not calculated 2.Text-based research rather than blood analysis
Ventilator-Associated Pneumonia	Ref. [52]	RF	Supervised	Accuracy 81%	1. Limited subjects 2. Slow performance
Appendicitis	Ref. [53]	RF	Supervised	AUC .98	1. Slow model performance
	Ref. [54]	DT	Supervised	Accuracy 94.69%	2. Small number of cases
	Ref. [55]	Gradient Boosting Tree	Supervised	Accuracy 95.31%	1. Small changes could cause model instability 2. Majority of samples were male
Urinary Tract Infection	Ref. [56]	ANN	Supervised/Unsupervised	Accuracy 98.30%	1. Works with complex data 2. Slow implementation

datasets (DASS42 and DASS21) were: NB, Bayes Network, k-star, Local Nearest Neighbor, MLP, Radial Basis Function Network (RBFN:NN-based), J48 (tree-based algorithm) and RF [61]. The results were analyzed using the Waikato Environment for Knowledge Analysis (WEKA) data mining tool. For testing and accuracy improvement, five-fold cross validation was implemented after training. The RBFN performed the best for anxiety in the DASS42 survey and the RF performed the best for the DASS21 survey (depression: RBFN and RBFN, stress: RBFN and MLP). Limitations of the latter include a small, unbalanced dataset.

2.5 COVID-19

Global life has been recently disrupted by the spread of COVID-19, and screening has not been very effective for determining the number of cases since the spread of the

virus is overtaking countries' capacity to test for it. In many parts of the world, the pandemic is too widespread to rely on these clinical tests, and many infected people do not display symptoms severe enough to go to a hospital for tests (81% as of April 1, 2020) [184]. In-person, clinical testing methods are proving inadequate for identifying cases because the availability, costs, risks to doctors and nurses, and likely event of a sick person not going to a hospital. There has been an attempt to use ML concepts to account for these problems by identifying COVID-19 cases without in-person tests. The process in this paper is to use an app to record audio of a person's cough and determine if it is "COVID-19 likely", "COVID-19 unlikely", or "Test inconclusive". This process is still early and has not been approved or widely accepted, but it represents a significant and important application of ML in DDRI. The data for this system was in the form of

TABLE 3: Table of common mental diseases and techniques for detection

Disease	Reference Number	Technique	ML Types	Measurement	Challenges (If available)
Schizophrenia	Ref. [57]	DCNN	Supervised	Accuracy 97%	1. Binary Labels 2. Cannot judge functional abnormalities- only structural
Bipolar Disorder	Ref. [58]	SVM	Supervised	Accuracy 80%	1. Prone to overfitting 2. Lower classification performance than other mental diseases
Panic Disorder	Ref. [59]	LR	Supervised	Accuracy 78.4%	1. Cross sectional design 2. Limited sample size
Major Depressive Disorder	Ref. [60]	RVM	Supervised	Accuracy 90%	1. Model should be extended 2. Comorbidity excluded
Anxiety	Ref. [61]	RF RBFN	Supervised Supervised or Unsuper- vised	Accuracy 100% Accuracy 97.48%	1. Imbalanced Data 2. Small sample size
Post-Traumatic Stress Disorder	Ref. [62]	SVM, NB, LR, & RF Ensemble	Supervised	AUC 0.85	1. Limited understanding of PTSD development 2. Limited sample size
Autism	Ref. [63]	SVM	Supervised	Accuracy 95%	1. Does not perform well with many features 2. Not suitable for large datasets
Obsessive Compulsive Disorder	Ref. [64]	RF	Supervised	Accuracy 70%	1. Slow prediction Speed 2. Low Accuracy
Attention Deficit Hyperactivity Disorder	Ref. [65]	DT-based (LightGBM)	Supervised	Accuracy 80%	1. ADHD is comorbid 2. Data contains noise

individual's cough samples gathered through an app and returned if they were likely to have COVID-19 [184]. Coughing is a common symptom of many diseases and is not unique to COVID-19, but research showed that coughs caused by different diseases produce identifiably different sounds. The process of this system was to allow users to record audio of their cough and have the model in the app produce its prediction about whether the cough was likely from COVID-19 or not. The design was constructed with the idea that it would be accessible to many people and could address the problems with the in-person testing method. The system for this app included three distinct classifiers that all had to agree in their conclusion of likely/unlikely for the test to give a prediction, otherwise it was inconclusive [184]. Since the app relied on cough recordings, the system had to begin a cough detection algorithm that was trained using an online database to differentiate the sound of the cough from any background noise. A DL Multi Class classifier, a classical ML based Multi Class classifier, and a DL Binary Class classifier were used in parallel to determine if a cough was likely to be COVID-19 or unlikely to be COVID-19. These classification models were implemented with a CNN, a SVM, and another CNN respectively.

The model showed promising results even though it is still new and does not yet have the volume of data that would be desired for this kind of system [184]. Overall, the model is reporting a true positive rate of 89% and a true negative rate of 81%. These are promising results especially since many of the cases that would have been misdiagnoses are caught as "inconclusive". In a breakdown of the three individual classifiers, the deep transfer learning-based multi-class classifier (DTL-MC) showed

the highest accuracy of the three (92.64%). The accuracy is expected to change as more data becomes available for the model to improve the DL classifiers. This system is still early in its development, but it shows promise and is a great example of how ML can be useful for disaster scenarios in the medical field.

2.6 MRSA

Methicillin-resistant *Staphylococcus aureus* (MRSA) is a serious infectious disease that is resistant to antibiotics and acquired primarily in the hospital setting [83]. It is important to have early detection of MRSA before it progresses and identifying risk factors for MRSA contraction can aid in the detection. Data from Electronic Health Records (EHRs) from check-in and throughout hospitalization was gathered to predict MRSA contraction risk. LR, SVM, and RF were used to detect text-based patterns in the data.

Data from the Medical Information Mart for Intensive Care III (MIMIC III) dataset was used for analysis. The populations observed were MRSA-positive and MRSA-negative. The dataset was unbalanced because a majority of the population was MRSA negative, so random sampling was implemented to create balanced groups of MRSA positive and negative patients. The researchers looked at the Microbiology Events Table to extract the test associated organism 80293 (MRSA) to obtain the diagnosis time for positive results.

Feature engineering was then carried out by separating features into two groups: 9 on-admission features and 80 throughout-stay features. On-admission features are unchanging, such as demographics, admission type, ad-

TABLE 4: Table of common infectious diseases and techniques for detection

Disease	Reference Number	Technique	ML Types	Measureme	Challenges (If available)
COVID-19	Ref. [66] [67]	SVM	Supervised	Accuracy 99.689% 95.33%	1. Prone to overfitting 2. Not suitable for large datasets
Pneumonia (general)	Ref. [68]	Linear Discriminant Analysis	Supervised	Accuracy 99.41%	1. Specific assumptions must be met 2. Diminisionality
Bacterial Meningitis	Ref. [69]	DT	Supervised	Accuracy 94.3%	1. Limited interpretability 2. Only 30 cases used for training 1. Data should be more diverse 2. Restrictive
	Ref. [70]	Bayesian Belief Network	Supervised/ Unsupervised	Precision 99.99%	
Tuberculosis	Ref. [71] [72]	SVM	Supervised	Accuracy 96.68% / .92%	1. Small sample size 2. Slow performance 1. Model interpretability 2. Works well with large sample size
	Ref. [73]	RF	Supervised	AUC .8646	
Malaria	Ref. [74]	CNN	Supervised/ Unsupervised	Accuracy 96.7%	1. Slow performance 2. Works well with complex Data
Dengue	Ref. [75]	ANN	Supervised/ Unsupervised	Accuracy N/A	1. Slow performance 2. Works well with complex data
Pertussis	Ref. [76]	DT	Supervised	Accuracy 79.9%	1. Possible recall bias 2. Low accuracy
Leprosy	Ref. [77]	RF	Supervised	Sensitivity / Specificity 93.97% 87.09%	1. Sample was not diverse 2. Interpretability
E. Coli	Ref. [78]	PCA & SVM	Supervised	N/A	1. Prone to overfitting 2. Not suitable for large datasets
Hepatitis	Ref. [79]	Feed-Forward NN	Supervised/ Unsupervised	Accuracy 98%	1. There must be major computational effort to avoid overfitting 2. Works well with large datasets 1. Not suitable for large datasets Slow performance
	Ref. [80]	SVM-SA	Supervised	Accuracy 96.25%	
Zika Virus	Ref. [81]	RF	Supervised	Accuracy 94.19%	1. Ideal for quantitative data 2. Training is slow
Influenza Virus	Ref. [82]	SVM	Supervised	Sensitivity 97.15%	1. Signs and symptoms should be reduced 2. Not suitable for large datasets
MRSA	Ref. [83]	RF	Supervised	AUC .95	1. Models should be expanded 2. Develop a way to enhance risk prediction
Bacterial Meningitis	Ref. [69]	DT	Supervised	Accuracy 94.3%	1. Limited interpretability 2. Only 30 cases used for training 1. Data should be more diverse 2. Restrictive
	Ref. [70]	Bayesian Belief Network	Supervised/ Unsupervised	Precision 99.99%	
Tuberculosis	Ref. [71] [72]	SVM	Supervised	Accuracy 96.68% / .92%	1. Small sample size 2. Slow performance 1. Model interpretability 2. Works well with large sample size
	Ref. [73]	RF	Supervised	AUC .8646	

mission location, and insurance. Throughout-stay features consist of laboratory results and vital signs.

For binary classification purposes, MRSA-positive was labeled as 1, and MRSA-negative was labeled as 0. LR, SVM, and RF classifiers were trained on 80% of the data and tested on 20%. Five-fold cross validation was used on the training data. Results showed that MRSA can be predicted effectively with acceptable accuracy (AUC < .95) days before it is diagnosed and has almost perfect prediction accuracy (AUC .98) one day before diagnosis using the RF classifier.

2.7 Brain Cancer

Diagnosing brain cancer is critical and determining methods of classifying brain neoplasms (abnormal growth) is important for early intervention and proper treatment selection [84]. This research focuses on glioma classification (meningioma, glioma grade II, grade III, grade IV, and metastasis), a type of brain tumor originating in the glial cells in the brain/spine, from MRI data. Ninety-seven patients were analyzed that had brain neoplasms.

Data preprocessing included noise reduction, bias-field correction, sequence co-registration, skull stripping, and histogram matching. The following region of interests (ROIs) were used for attribute extraction:

- Neoplastic, enhancing
- Neoplastic, non-enhancing

- Necrotic
- Edematous

Then, 152 features were used for analysis (i.e., age, image intensity, tumor shape characteristics, etc.). Pearson's correlation was used to standardize the variables. The following classifiers were evaluated on the data using a leave-one-out strategy: J48 Tree, KNN, Voting Feature Intervals (VFI), SVMs, and NB. The best performance was achieved by the KNN classifier using the best first search method and wrapper evaluator with 96.3% accuracy.

2.8 Gastric Cancer

Gastric cancer is an intense type of cancer where the survival rate increases by up to 62.8% if diagnosed early (if the cancer is advanced, the survival rate is only about 24%) [87]. Raman spectra of amino acids from samples of saliva were used for binary classification to determine if a patient had gastric cancer or was healthy. Fourteen Raman peaks (intensity and wavelength). Twenty patients with early gastric cancer were sampled, eighty-four with advanced gastric cancer, and 116 healthy controls. The data preprocessing steps included baseline correction using the fourth order polynomial and noise removal using the second-order Savitzky-Golay filter.

The feature extraction process required calculating the area under each Raman peak (14) which shows information about different organic compounds in the 2D saliva

TABLE 5: Table of common non-infectious diseases and techniques for detection

Disease	Reference Number	Technique	ML Types	Measurement	Challenges (If available)
Brain Cancer	Ref. [84]	kNN	Unsupervised	Accuracy 96.9%	1. Slow prediction with large data 2. Computationally expensive
Hepatocellular Carcinoma	Ref. [85]	SVM	Supervised	F1-Score 0.8849	1. Not robust to outliers 2. Not suitable for large data
Gastric Cancer	Ref. [86]	GBDT	Supervised	AUC 91%	1. Small max-depth parameter
	Ref. [87]	SVM	Supervised	Accuracy 97.18%	2. Works better with less features 1. Not suitable for large datasets
	Ref. [88]	DCNN	Supervised Unsupervised	Sensitivity near 100%	2. Slow performance 1. Works better with more features 2. Works well with large data size
Kidney Cancer	Ref [89]	DCNN (KidNet)	Supervised/ Unsupervised	Accuracy 86.1%	1. Slower performance 2. Works well with large sample sizes
Oropharyngeal Carcinoma	Ref [90]	Naive-Bayes	Supervised	Accuracy 65.9%	1. Not ideal for diseases that do not include multiple features 2. Accuracy should be improved
Prostate Cancer	Ref [91]	CNN	Supervised/ Unsupervised	Accuracy 85.4%	1. Small sample size 2. Need for external validation
Pancreatic Cancer	Ref [92]	SVM	Supervised	Accuracy 100%	1. Small sample size 2. Need for external validation

TABLE 6: Table of common deficiency diseases and techniques for detection

Disease	Reference Number	Technique	ML Types	Measurement	Challenges (If available)
Phenylketonuria (PKU)	Ref. [93]	LR	Supervised	Sensitivity 96.8%	1. Linearity Assumption in LR 2. More research necessary
MCADD	Ref. [93]	LR	Supervised	Sensitivity 96.8%	1. Linearity Assumption in LR 2. More research necessary

data. The data was entered into a supervised RBF kernel-based SVM classifier which was tested using 10-fold cross validation. For optimization, the epsilon value and cost function was used. An accuracy of 97.13% was achieved by the model. The benefits of this model are its cost-effectiveness, non-invasiveness, and quick performance.

2.9 PKU/MCADD

PKU is a deficiency disease of phenylalanine, an amino acid that occurs in infants and can lead to seizures, behavior problems, etc. if not treated. MCADD is a deficiency disease that prevents the body from properly breaking down fat to use as energy [93]. The research used modern tandem mass spectrometry (MS/MS) of blood samples which can analyze multiple compounds with high-throughput. The data samples were saved into a database and analyzed using ML classification algorithms. The following algorithms were used on the dataset: Discriminant Analysis, LR, DT, KNN, ANN, and SVM.

For feature selection, correlation-based techniques were used: gain ratio and relief to reduce data dimensions. For both deficiencies, LR achieved the best classification with a sensitivity > 96.8% and specificity > 99.98%.

2.10 Heart Disease

Increasing cardiovascular disease rates prompted [13] to perform a study aiming to determine the best ML methods to improve the detection ability of cardiovascular

issues. They used two public datasets on cardiovascular disease and compared accuracy and computation time of different ML algorithms to model the viability of using ML for prediction. The accuracy and time of these algorithms were also compared with the data subdivided by other categories (age, gender, education) to find potential increases in model performance. They found that the ML algorithms could consistently give high accuracy, and some had very fast model creation times, making them sustainable in detecting risks of heart disease. The techniques seemed showed an interchange between model creation time and accuracy. The study chose to use existing data on cardiovascular health from online data sets instead of testing cardiovascular sensors simultaneously. For the study, the NHANES [186] and Framingham Heart Study [187] datasets were used for the predictive model. The NHANES dataset was split into two separate files and subdivided data by categories. The unified form of the data was used as the final dataset for the ML models.

A variety of ML techniques were tested to compare the results according to the accuracy of classification and the time required to build the model. Table 13 compares the algorithms for each of the two used data sets.

Both datasets show a trade-off between speed and accuracy; a model that can make more accurate predictions will take longer to build. Each model was also tested while grouping the data according to different features of the subjects in question such as age, gender, and education

TABLE 7: Table of common inherited diseases and techniques for detection

Disease	Reference Number	Technique	ML Types	Measureme	Challenges (If available)
Breast Cancer	Ref. [94]	SVM-RBF	Supervised	Accuracy 98.26%	1. Limited Cancer Data 2. Difficult to train on large datasets. 1. Complexity 2. Difficult to train on large datasets.
	Ref. [95]	ANN	Supervised/ Unsupervised	Accuracy 98.57%	
Ovarian Cancer	Ref. [96]	CNN	Supervised	Accuracy 85%	1. Complexity 2. Slow performance 1. Difficult to train on large datasets. 2. Prone to overfitting
	Ref. [97]	RF & SVM	Supervised/ Unsupervised	Sensitivity 0.96	
Prostate Cancer	Ref. [98]	ANN	Supervised/ Unsupervised	AUC .9755	1. More ML techniques should be applied on same data 2. Complexity 1. High variability and small sample size of the data 2. Manually updating features using more advanced models will increase performance
	Ref. [99]	Xmas-net (CNN-based)	Supervised/ Unsupervised	AUC 0.84	
Colorectal Cancer	Ref. [100]	CNN	Supervised/ Unsupervised	Accuracy 81.2%	1. Complexity 2. Slow performance
Cystic Fibrosis	Ref. [101]	Gradient Boosted DT	Supervised	Accuracy 98%	1. Prone to overfitting 2. Computationally expensive
Leukemia	Ref. [102]	kNN	Unsupervised	Accuracy 98.38%	1. Requires large memory to store data 2. Processing can be expensive 1. Categorization of patient risk 2. Sample size and absence of strong predictors
	Ref. [103]	RF	Supervised	Accuracy 82.9%	
Sickle Cell Anemia	Ref. [104]	MLP & Logit-Boost	Supervised	Accuracy 99.5984%	1. Intelligent support and ML models should be combined to aid physicians in diagnosis 2. Excessive parameters
Skin Disease	Ref. [105]	NN	Supervised/ Unsupervised	Accuracy 96% - 98%	1. Complexity 2. Slow performance 1. Difficult to train on large datasets 2. Works well with small data 1. Complexity 2. Slow performance
	Ref. [105]	SVM	Supervised	Accuracy 91.66%	
	Ref. [106]	CNN & SVM	Supervised/ Unsupervised	Accuracy 100%	
Heart Disease	Ref. [79]	SVM	Supervised	Accuracy 94.60%	1. Does not handle multiple features well 2. Works well with small data samples
Diabetes	Ref. [79]	Naive-Bayes	Supervised	Accuracy 95%	1. Not idea for diseases that don't include multiple features 2. Assumes independent features 1. Not ideal for large datasets. 2. Complexity
	Ref. [21]	SVM	Supervised	Accuracy N/A	
Liver Disease	Ref. [79]	Functional Trees (FT)	Supervised	Accuracy 98.10%	1. Involves complex equations 2. Longer training period
Macular Disease	Ref. [107]	CNN	Supervised/ Unsupervised	Accuracy 85%	1. Small sample size 2. Accuracy may not be greater than professional diagnosis 3. Complexity
Down Syndrome	Ref. [108]	SVM	Supervised	Accuracy 97.92%	1. Slower training period 2. Not suitable for large data
Asthma	Ref. [39]	RF	Supervised	Precision 80.3%	1. Future research should be done using non-invasive data (i.e., habits, air pollution, etc.) 2. Difficult to interpret
Fragile-X	Ref. [109]	RF	Supervised	F1 Score 0.81	1. Only females with children who have FX were investigated 2. Speech impediments used for evaluation are not limited to FX
Thalassemia	Ref. [110]	RF	Supervised	Recall 100%	1. Difficult to interpret 2. Susceptible to overfitting 1. Susceptible to additional noise 2. Does not do well with high-dimensional data
	Ref. [111]	Synthetic Minority Over-Sampling Technique SMOTE	Unsupervised	Specificity 99.47%	
Tourette Syndrome	Ref. [112]	SVM-RFE	Supervised	Accuracy 89.13%	1. Feature selection takes a long time 2. Prone to overfitting
Von Willebrand Factor	Ref. [113]	RF	Supervised	Accuracy 86% - 94%	1. There is a slight issue of collinearity 2. Model interpretability

level. This also showed similar results to the previous tests.

The study employed mean squared error (MSE) as the loss function of choice and used varying sizes of data inputs and batch sizes for future-proofing purposes. Adequate results were achieved overall for cardiovascular risk prediction [188].

2.11 Diabetes

Various ML techniques were implemented against the Pima Indian Diabetes Dataset that contains individuals being tested for diabetes [183]. The study investigated the accuracy of algorithms such as DTs, SVMs, and kNNs on this database. The study, performed in MATLAB, found that algorithms such as LR, SVMs, and discriminant analyses worked best on this dataset, while variations of SVM (CG-SVM) and DT (fine tree) performed the worst.

2.12 Parkinson's Disease

The Gait Abnormality Detection System is an application of DL techniques using wearable sensors as an input [182].

The system incorporates various pressure and motion-sensing materials connected to information-gathering microcontrollers located on the subject being tested. The data from the sensors is then analyzed to model the gait of a typical human subject. The features gathered from this data are used to determine whether an individual is walking in a pattern similar to that of a patient suffering from various types of human gait disorders such as Parkinson's, Hemiplegia, or sensory Ataxia. The results showed findings that were accurate in testing but appeared to only identify a single class.

The results were indicative of successful gait abnormality detection. Multiple architectures and designs were tested to compare performance and the LSTM model showed the smallest error of .02 with less than 30 trials (epochs) for detecting healthy gate. While the accuracy for detection of sensory ataxic gaits was only 44.4%, the detection of Parkinson's achieved a 97.5% with the stacked LSTM model.

TABLE 8: Table of common degenerative diseases and techniques for detection

Disease	Reference Number	Technique	ML Types	Measurement	Challenges (If available)
Amyotrophic Lateral Sclerosis	Ref. [114]	RF	Supervised	AUC .96	1. Difficult to interpret 2. Susceptible to overfitting without proper hyperparameter tuning
	Ref. [115]	DT	Supervised	Accuracy 96.6%	
Alzheimer's Disease	Ref. [116]	DL & SVM	Supervised/Unsupervised	Accuracy 87%	1. Difficult disease to diagnose early-on 2. Accuracy is lower than preferred
Multiple Sclerosis	Ref. [117]	RF & SVM	Supervised	Accuracy 85.7%	1. Works well with complex features 2. Slow performance
Muscular Dystrophy	Ref. [118]	CNN	Supervised/Unsupervised	Accuracy 91.7%	1. Complexity 2. Slower performance
Rheumatoid Arthritis	Ref. [119]	SVM	Supervised	Accuracy 81.4%	1. Not suitable for large data 2. Complexity 1. More research on gene analysis for RA detection is necessary to increase significance 2. Interpretability
	Ref. [120]	LR & RF	Supervised	AUC significant	
Huntington's Disease	Ref. [121]	SVM	Supervised	Accuracy 97.14%	1. Takes longer time periods to train large data 2. Does not handle noise well 1. Unstable 2. Issues with accuracy of DTs
	Ref. [122]	SVM & DT	Supervised	Accuracy 100.0%	
Friedreich Ataxia	Ref. [123]	RF Regression	Supervised	RHO 0.73	1. Repeated measures should be performed 2. Random effects should be incorporated
Marfan Syndrome	Ref. [124]	DT	Supervised	AUC 0.97	1. Prone to overfitting 2. Gives low prediction accuracy
Thymoma-associated myasthenia gravis	Ref. [125]	3D-DenseNet-DL	Supervised/Unsupervised	AUC 0.734	1. Unavoidable selection bias 2. Absence of serum AChR binding antibodies status
Parkinson's Disease	Ref. [126]	ANN & Levenberg-Marquardt	Supervised/Unsupervised	Accuracy 95.89%	1. Unexplained network functioning 2. Sensitive to noise
Polio	Ref. [127]	k-Means	Unsupervised	N/A	1. Precision 2. Not much related work available
Osteoporosis	Ref. [128]	RF	Supervised	Accuracy 93%	1. More research necessary 2. Performance
Macular Degeneration	Ref. [129]	CNN (Trained on Multiple Modalities)	Supervised/Unsupervised	Accuracy 96%	1. Limited sample size 2. Limited generalizability

2.13 Alzheimer's Disease

Alzheimer's is a degenerative disease effecting mental function, primarily in older adults with main symptoms consisting of confusion and loss of memory. Multimodal Deep learning was used to analyze image-based MRI data, genetic single nucleotide polymorphisms (SNPs), and EHR data to characterize patients as having Alzheimer's Disease (AD), Mild Cognitive Disorders (MCI), or Controls (CN) [116]. For genetic and clinical data feature extraction, SDAEs were used. For the image-based data, 3D-CNNs were used.

Preprocessing of the MRI data included noise filtering, skull stripping, brain tissue segmentation, and normalization and co-registration of images. Then images from the right amygdala, left and right angular, left and right cerebellum, left and right Hippocampus, left and right occipital regions, and left and right superior temporal regions were analyzed. Preprocessing of the EHR data included extracting 1680 common features, normalizing quantitative data (1-2 range), and assign binary data values of 1 and 2. Preprocessing of the genetic data includes filtering and feature selection to discard items with low genotype quality and allele frequency, high per-site

missing rate, and significant Hardy-Weinberg equilibrium p-value. Then genes were selected that contained SNPs using minimum redundancy maximum relevance.

DL models were found to have better performance than classical or "shallow" ML (i.e., KNN, RF, DT, SVM.) The results showed that for Imaging, EHR, and SNP combined, DL and SVM achieved the highest accuracy of 87%. EHR and SNP combined showed an accuracy of 87% with DL and SVM. For Imaging and EHR data combined the highest accuracy was 88% from DL and SVM. Imaging and EHR combined showed the highest accuracy of 78% from "shallow" models which could be attributed to small sample size. The main limitations of the study were the size of the data and the lower accuracy. Future research should perform testing on larger datasets.

2.14 HIV

Human immunodeficiency virus (HIV-1) is the most dangerous social disease and can lead to the incurable, life-threatening acquired immunodeficiency syndrome (AIDS). Kameris, a ML software package that is alignment-free and works on k-mer frequencies was developed for this study and is an effective subtype classifi-

TABLE 9: Table of common social diseases and techniques for detection

Disease	Reference Number	Technique	ML Types	Measurement	Challenges (If available)
HIV	Ref. [130]	SVM	Supervised	Accuracy 96.49%	1. Not suitable for large data 2. Not suitable for noisy data
Gonorehea	Ref. [131]	Bayesian Additive Regression Trees	Supervised	AUC 0.75	1. More research needed 2. Poor balance between sensitivity and positive predictive value
Syphilis	Ref. [131]	Bayesian Additive Regression Trees	Supervised	AUC 0.75	1. More research needed 2. Poor balance between sensitivity and positive predictive value
Chlamydia	Ref. [131]	Bayesian Additive Regression Trees	Supervised	AUC 0.75	1. More research needed 2. Poor balance between sensitivity and positive predictive value
Human Papillomavirus/ Oropharyngeal Squamous Cell Carcinomas	Ref. [132]	Deep Learning	Supervised/ Unsupervised	AUC 0.74	1. Minimal interrater reliability 2. DL requires large data size
Hepatitis B	Ref. [133]	SVM	Supervised	Accuracy 87.7%	1. Not suitable for noisy data 2. Not suitable for large data
Trichomoniasis	Ref. [134]	RF	Supervised	AUC (women) 0.87 AUC (men) 0.83	1. Urine sediment was examined. Ncleic acid-based test may achieve different results 2. Sample was obtained from reference hospital

TABLE 10: Table of common self-inflicted diseases and techniques for detection

Disease	Reference Number	Technique	ML Types	Measurement	Challenges (If available)
Lung Cancer	Ref. [135] [136]	SVM	Supervised	Accuracy 92%	1. Proper pre-processing needed for higher accuracy 2. False objects should be eliminated. 1. Works better with more features. 2. Complexity 1. Computational cost 2. Joint probability estimation
	Ref. [137] [138]	ANN	Supervised/ Unsupervised	Sensitivity / Accuracy 95.9%/96.67%	
	Ref. [139]	DBN	Supervised/ Unsupervised	Accuracy 81.19%	
Obesity	Ref. [140]	Generalized Linear Model, RF, & Partial Least Squares	Supervised	Accuracy 89.68%	1. RF- model Interpretability 2. GLM- sensitive to outliers
Alcoholism	Ref. [141]	ADT	Supervised/ Unsupervised	Accuracy 86.1%	1. Mainly male subjects 2. Imbalanced data
Anorexia	Ref. [142]	SVM Ensemble	Supervised	F1-Score 85%	1. 8 Neural submodels within the ensemble should be tested separately to improve the final model 2. Does not include linguistic features and metadata
Binge-Eating Disorder	Ref. [143]	RF	Supervised	Accuracy 81.25%	1. Data only contained healthy individuals 2. Medication/Personality could have an influence on EEG data used

cation method for HIV-1 [130]. Alignment-based methods, where the input is aligned with a reference subtype, are computationally expensive, require many ad-hoc settings, and do not have adequate performance on highly divergent genome regions. An alignment-free-based method was recommended to address the drawbacks.

K-mer (substrings of length k) frequency vectors were used for supervised training. Actual HIV-1 datasets were used for testing the model. The most accurate performance was achieved from SVM classifiers, multilayer perceptron, and logistic regression. A frequency length of k=6 proved to achieve the best combination of accuracy and speed. The Kameris ML software package achieved a 97% classification accuracy.

2.15 HPV

With the evolution of ML, the detection of molecular and cellular changes through images is becoming more accurate, which can aid in the detection of oropharyngeal squamous cell carcinoma (OPSCC) in patients with human papillomavirus (HPV) [132]. HPV is the most common social disease and can lead to OPSCC, which is critical to be diagnosed early. OPSCC is often diagnosed in later stages which leads to high mortality. A DL model evaluated hematoxylin and eosin (H&E) stain images to predict HPV association in patients with OPSCC with an AUC of 0.74. The main limitation of this study was interrater reliability.

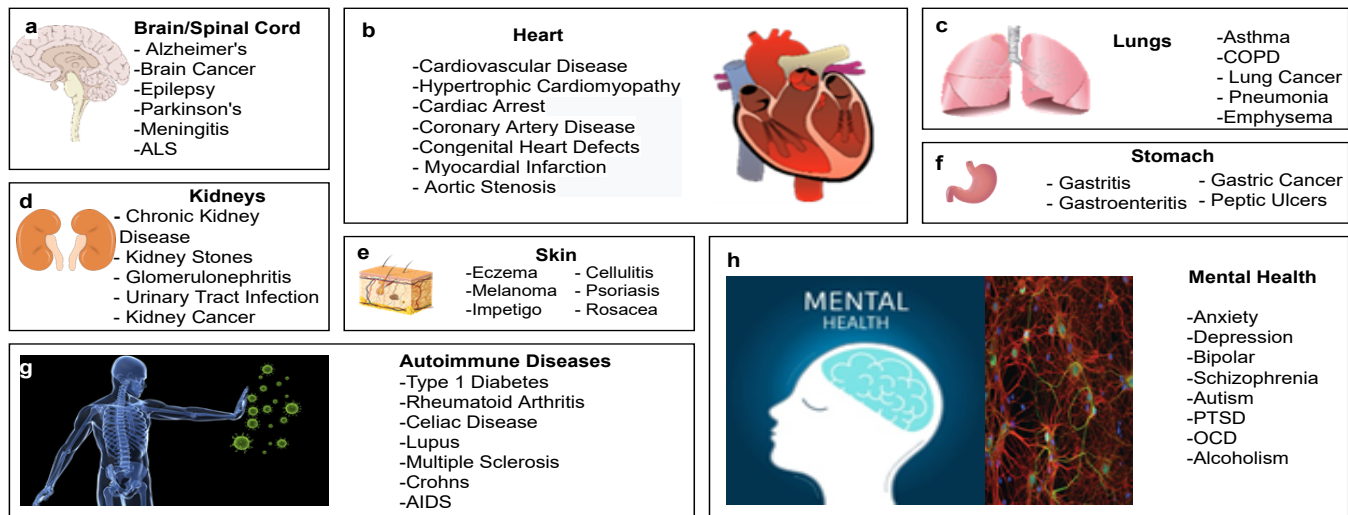


Fig. 5: Some parts of the body and common associated diseases/conditions: **a.** Brain [144] diseases are detected using diagnostic tests such as a balance test, biopsy, blood test, Electroencephalography (EEG), Computed Tomography (CT) scan, gait mapping, MRI, ultrasound, etc. [145] **b.** Heart [146] diseases are typically diagnosed using diagnostic tests such as ECT/EKG, Echocardiogram, stress test, CT scan, or MRI [147]. **c.** Lungs [148] diseases are detected using diagnostic tests such as bronchoscopy, CT scan, Endobronchial Ultrasound (EBUS), Exhaled Nitric Oxide Test, Lobectomy, etc. [149] **d.** Kidney [150] diseases are detected using blood tests, urine tests, imaging tests, and kidney biopsies [151]. **e.** Skin [152] diseases are diagnosed by skin biopsies, visual exams by a physician, and ultraviolet light for pigment abnormality detection (Wood's Lamp Examination) [153] **f.** Stomach [154] diseases are diagnosed using an endoscopy, x-ray, upper gastrointestinal series (UGI), ultrasound, etc. [155]. **g.** Autoimmune Diseases [156] are difficult to diagnose. Physicians look at combinations in symptoms and blood markers, possibly even a biopsy to aid in detection of these diseases [157]. **h.** Mental Health [158] disorders are diagnosed by psychological evaluations that often rely heavily on questionnaires, physical examinations to rule out physical issues, and lab tests for drug screening or thyroid problems that could be causing mental instabilities [159].

2.16 Obesity

Obesity is a self-inflicted or inherited disease, affecting around 42.5% of Americans in 2017-2018 [189]. An individual is labeled as obese when their body mass index (BMI) is greater than 30 [190]. Obesity is a risky disease because it can lead to heart disease, type II diabetes, cancer, sleep apnea, severe COVID-19 symptoms, etc. Obesity can be categorized into different levels based on BMI: 1 (30.0–34.9), 2 (35.0–39.9), and 3 (more than 40).

An ensemble model composed of a generalized linear model, RF, and partial least squares can predict obesity levels with 89.68% accuracy [140]. Feature selection was not used for small to medium datasets, which may improve the model with future work. The model was ensemble in R using Python as the interface. The ensemble approach works by importing the dataset, building the model, receiving the user input, and if the data is sufficient then obesity is calculated and if it is not adequate, it is discarded.

2.17 Lung Cancer

Lung cancer is often a self-inflicted disease where cigarette smoking is the number one risk factor associated with

80% to 90% of lung cancer deaths [191]. An ANN model was constructed to detect the presence of lung cancer using a text-based dataset [137]. For data preprocessing and transformation, 14 input attributes were used for prediction, certain variables were normalized (gender, age, lung cancer) where the binary variables were assigned the values 0 and 1 (gender, and lung cancer).

There were 1,418,105 training cycles where the training error rate is < 1% and the prediction accuracy was 96.67% with the ANN model. Out of all of the variables, age was found to have the largest effect on the presence of lung cancer.

3 DATASETS

In this section, we discuss some of the most commonly used benchmark datasets for the evaluation of ML-based solutions for the diagnosis of different diseases.

3.1 Epilepsy Dataset

The literature reports several interesting ML-based solutions for epileptic seizure recognition. The following are

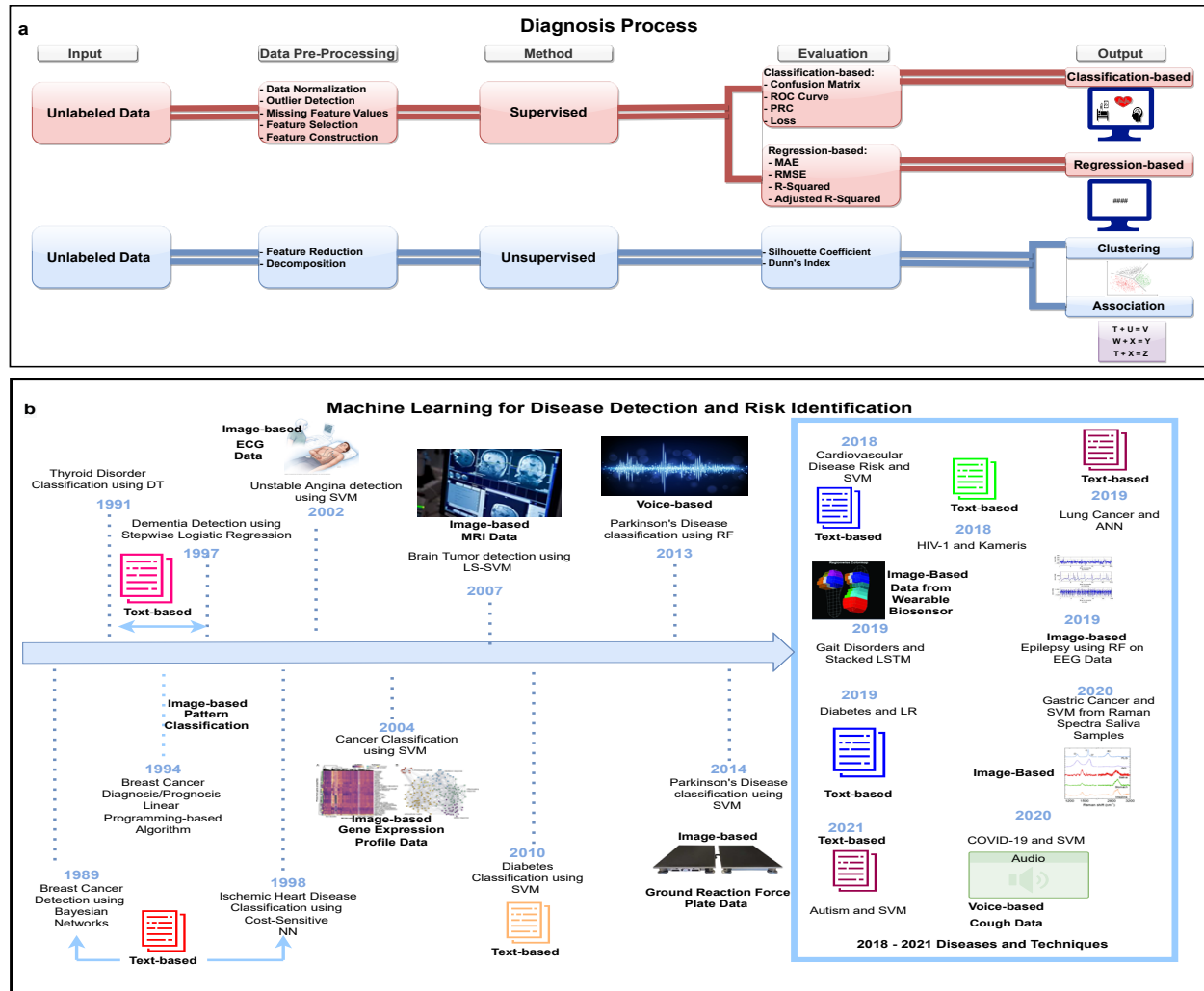


Fig. 6: **a.** Representation of the diagnosis process with the use of ML models. Labeled data is preprocessed [160] and input into either a Regression-based model or a Classification-based supervised ML model. The output for a Regression-based model is a continuous number. The output for a Classification-based model is a categorical classification [161] [162] [163] [164]. Unlabeled data is preprocessed and input into an unsupervised ML model. The most common unsupervised ML models are clustering and association [165]. **b.** Representation of a timeline of various common diseases and the ML techniques used for detection from 1989 to 2021. First, Bayesian Networks are used to detect breast cancer from a dataset used to classify malignant versus not malignant [166]. In 1991, thyroid disorders were used a DSS from patient record data and a DT were used to classify Thyroid disorders from routine cases [167]. In 1994, breast cancer diagnosis/prognosis was determined using Linear Programming-based ML [168]. In 1997, Dementia was detected using Stepwise Logistic Regression on mental performance data (questionnaire, exam, and task) obtained from healthy and mildly demented individuals [169]. In 1998, Ischemic Heart Disease was classified using cost-sensitive NNs from a dataset containing patients' diagnostic procedures [170]. In 2002, unstable angina was detected using GAs in combination with SVM from data collected from hospital patients [171] [172]. In 2004, cancer was classified (colon, lung, and prostate) using gene expression profiles and applying SVM along with active learning [173] [174]. In 2007, brain tumors were detected using an LS-SVM classifier on image slices of MRI data identifying abnormal brain volume [175] [176]. In 2010, diabetes was classified using a SVM on the NHANES dataset [177]. In 2013, Parkinson's disease was classified from voice data using a RF classifier [178] [179]. In 2014, Parkinson's disease was determined using gait classification data from a force plate measuring Ground Reaction Force using SVM classifier [180] [181]. Some of the main topics discussed in this paper occurred from 2018 to 2021 and are featured in the rectangular portion located on the right side of the figure. These topics include: Cardiovascular disease [13], HIV-1 [130], Lung Cancer [137], Gait Disorders (i.e., Parkinson's) [182], Epilepsy [40], Diabetes [183], Gastric Cancer [87], Autism [63] and COVID-19 [184].

TABLE 11: Table of the results of multiple ML techniques on various heart diseases for classification [31]

Heart Disease	Accuracy	Algorithm
Myocardial Infarction	97%	ANN
Myocardial Perfusion Scintigraphy	87.5%	ANN
Heart Failure	87.6%	CART
Heart Valve [185]	97.4%	Neural Network Ensemble
Arrhythmia Screening	95.6%	SVM
Acute Coronary Syndrome	72%	Logistic Regression
Coronary Artery Disease	92.5%	Artificial Immune Recognition
Chest-Pain	91%	Genetic Algorithms & Multi-Criteria Decision Analysis

some of the most commonly used datasets used in literature for the training and evaluation of these solutions.

- **Epileptic Seizure Recognition Data Set [192]:** The dataset aims at epileptic seizure recognition in electroencephalogram (EEG) recordings. The data is collected from 500 individuals each having 4097 data points recorded for 23.5 seconds. The dataset is composed of five different categories including (i) data recorded with closed eyes, (ii) data recorded with opened eyes, (iii) region of tumor is identified, (iv) recording from the region of the tumor, and (v) recording of seizure activity, which represent the individual having an epileptic seizure. Each recording is composed of 178 different readings (i.e., features).
- **EEG Dataset of Epileptic Seizure Patients [193]:** The dataset is composed of EEG recordings of a total of 40 epileptic seizure patients including both male and female patients in the of 4 to 80 years of age. The data is recorded through the Allengers VIRGO EEG machine considering the International standard and procedures. The data is recorded from 16 channels at the rate of 256 samples per second.
- **Neonatal EEG Recordings Dataset [194]:** The dataset is composed of EEG recordings from human neonates each lasted approximately one hour recorded through a NicOne EEG amplifier. The dataset is manually annotated by the domain experts where each sample is annotated by at least three different experts, and the final label is assigned on the majority voting basis. All the data

samples are divided into two categories one representing samples with seizure and other non-seizure samples.

3.2 Chronic Kidney Disease Dataset

The following are some of the most commonly used benchmark datasets for chronic kidney diseases detection.

- **Kidney Disease Dataset [195]:** This dataset aims at chronic kidney diseases detection, and is collected over a period of two months. The dataset covers a total of 24 different features including 11 numeric and 13 nominal features. All the data samples are annotated either having chronic kidney disease or not having the disease. Some key features include *age, blood pressure, sugar, red blood cells, pus cell, blood urea, hemoglobin, white and red blood cell count, hypertension, and diabetes mellitus*.
- **CKKD [196]:** The dataset provides the details (both personal information as well as laboratory investigation) of kidney patients and aims to provide useful information for quantitative renal physiology. The dataset covers a wide range of information including demographics, medical history, laboratory test results, vital signs, and personal statistics/details, such as age and weight. The data is collected at the Department of Nephrology, PGIMER, Chandigarh from around 14,024 patients. However, 3,823 samples are removed due to missing information leaving a total of 10201 patient records only. The dataset is annotated for six different types of diseases including (i) Diabetic nephropathy, (ii) Chronic glomerulonephritis, (iii) Congenital disease, (iv) Cystic disease, (v) Obstructive uropathy, and (vi) Heredofamilial.

3.3 Autism Dataset

The following are some of the most commonly used autism datasets used in the literature.

- **Autism Screening Adult Data Set [197]:** The dataset aims at the classification and prediction of autistic spectrum disorder in adults. In total, the dataset provides 20 different types of features including 10 behavioral features. The rest of the 10 features represent individuals' characteristics that have proved to be effective in detecting autistic spectrum disorder. Moreover, the dataset is composed of 704 different samples.
- **DREAM Dataset [198]:** The dataset provides a collection of behavioral data recordings collected from 61 children diagnosed with autism spectrum disorder during robot enhanced therapy. A total of 3000 therapy sessions with more than 300 hours of therapy are conducted for the collection of the dataset. Some of the sessions are conducted by

TABLE 12: Summary of existing surveys and reviews related to Disease Detection and Risk Identification

Algorithm	NHANES Dataset		Framingham Dataset	
	Time (seconds)	% Accuracy	Time (seconds)	& Accuracy
Naive-Bayes	7.20	95.7%	28.90	89.9%
Decision Tree	56.7	97.6%	77.40	90%
Logistic Regression	1.71	96.4%	16.17	90%
K-Nearest Neighbors	2.70	80.8%	88.80	90.1%
Random Forest	214	98.5%	2180	90.1%
Bagging	3.50	96.5%	N/A	N/A
Support Vector Machine	N/A	N/A	468	90.2%
Ensemble	N/A	N/A	15.10	89.3%
Neural Network	N/A	N/A	200	89%

TABLE 13: Performance of machine-learning techniques on both the NHANES and Framingham datasets [13]. The NHANES dataset (left) shows LR performing the fastest at 1.71 secs with 96.4% accuracy, and the RF showing the greatest accuracy at 98.5%, with a training time of 214 secs. The Framingham dataset (right) shows ensemble tree method performing the fastest at 15.10 secs with 89.3% accuracy, and SVMs being the most accurate with 90.2% accuracy with a training time of 468 secs.

recording children's interactions with a social robot namely NAO under the supervision of a therapist. In the rest of the sessions, the children directly interacted with a therapist. The key features (i.e., factors recorded) of the dataset include *body motion*, *head position and orientation*, and *eye gaze variables*.

- **Urinary Element Compositions Dataset [199]:** The dataset aims at analyzing the association between urinary essential elements of the mothers of the children with autism spectrum disorder and typically developing children. In total, data is collected from 28 mothers having children with autism spectrum disorder and 29 mothers having typically developing children. We note that in the study mothers of children in the range of 2 to 5 years of age.
- **Autism Brain Imaging Data Exchange [200]:** The Autism Brain Imaging Data Exchange (ABIDE) is an international initiative to collect functional and structural brain imaging data across the world to facilitate the research on autism spectrum disorder. More than 24 laboratories worldwide are part of the initiative. As a result, several data collection of autism brain imaging, such as ABIDE I and ABIDE II, are provided.

3.4 Anxiety Dataset

Some publicly available datasets for ML-based anxiety detection are:

- **Anxiety Dataset [61]:** The dataset aims to facilitate the automatic analysis and assessment of depression and anxiety using ML techniques. To this aim, the dataset provides a total of 39,776 samples is collected using a survey composed of 42 different questions having an equal number of questions for depression, anxiety, and stress. All the samples have annotated either depression, anxiety, or stress. Moreover, the severity levels of these psychological disorders.
- **Depresjon Dataset [201]:** The dataset could be used for multiple applications including depression type classification, Montgomery-Asberg Depression Rating Scale (MADRS) score prediction, and sleep pattern analysis. The dataset provides the actigraph data, which is collected over a period of time, for each instance. The key features covered by the dataset include *timestamp*, *date*, *activity*, and *MADRS scores*. The MADRS scores include several readings, such as patient ID, number of days of measurements, gender, age, afftype (i.e., bipolar, unipolar depressive, or bipolar I), melanch, type of patient (i.e., inpatient or outpatient), education of the patient, marital status, job, etc.

3.5 COVID-19 Dataset

Some publicly available dataset in the domain include:

- **Coswara dataset [202]** The dataset by Indian Institute of Science (IISc) Bangalore to build diagnostic

TABLE 14: Table of results from the diabetes study showing accuracy of common ML algorithms [183]. The value of the most accurate technique, LR, is bolded at 77.9%.

ML Techniques	Classifier	Avg. Accuracy
DT	FT	69.0%
	MT	73.7%
	CT	75.3%
DA	LD	77.5%
	QD	74.1%
LR	LR	77.9%
SVM	L-SVM	75.5%
	Q-SVM	74.7%
	C-SVM	77.6%
	FG-SVM	75.1%
	MG-SVM	72.8%
	CG-SVM	65.5%
k-NN	Fine k-NN	76.0%
	Medium k-NN	73.2%
	Coarse k-NN	73.4%
	Cosine k-NN	74.5%
	Weighted k-NN	72.8%
	cubic k-NN	74.3%
Ensemble Learners	Boosted Trees	75.0%
	Bagged Trees	73.2%
	Subspace Discriminant	76.6%
	Subspace k-NN	71.6%
	RUS Boosted Trees	73.4%

tool for COVID-19 detection. The dataset provides the audio recordings such as breathing, cough and speech sounds of an individual. The volunteers from across the globe, age groups and health conditions were encouraged to record their sound data in a quiet environment using a web connected device. The volunteers record their acoustic data corresponding to 9 audio categories, namely, (a) shallow and deep breathing (2 types), (b) shallow and heavy cough (2 types), (c) sustained phonation of vowels [æ] (as in bat), [i] (as in beet), and [u] (as in boot) (3 types), and (d) fast and normal pace number counting (2 types). The whole process takes 5-7 minutes.

- **The Ovation COVID-19 dataset [203]** contains millions of deidentified diagnostic test results collected in real time (hour by hour) from testing facilities outside a normal hospital setting. This dataset includes normalized patient metadata, ICD10 diagnosis codes, qPCR test results, including assay-level CT values (a measure of the viral load of the SARS-CoV-2 virus), instrument/test kit details, and detailed dates with which to calculate test turnaround times. As COVID-19 testing scales up

across the country, this data continues to represent 2-3 % of all testing in the US each week.

- **Mendeley dataset [204]** consists of Non-COVID and COVID cases of both X-ray and CT images. The associated dataset is augmented with different augmentation techniques to generate about 17099 X-ray and CT images. The dataset includes 5500 Non-COVID images and 4044 COVID images, 2628 Non-COVID images and 5427 COVID images.

3.6 MRSA Dataset

No publicly available dataset found

3.7 Brain Cancer Dataset

The following are some commonly used datasets for the evaluation of brain cancer detection algorithms.

- **BRATS datasets [205], [206]:** These datasets are provided as part of a benchmark challenge namely Multimodal Brain Tumor Segmentation Challenge. Several versions of the dataset are available as part of the competition for several years. For instance, BRATS 2012 dataset provides a collection of patient and synthetic images. The patient data is annotated using 4 different labels, namely (i) necrosis, (ii) edema, (iii) non-enhancing tumor, (iv) enhancing tumor. The synthetic images are annotated using two different labels, namely (i) edema or (ii) tumor core. The BRATS 2018 dataset, which is one of the mostly commonly used dataset in the literature, provides a collection of multi-contrast MR scans acquired at four different centers using a diversified set of MR scanners. In total, 65 different scans included both low-grade and high-grade glioma patients are provided in the dataset. The annotations are provided by domain experts by manually analyzing and identifying the location and structure of the tumors.
- **Cancer Imaging Archive (TCIA) Brain Tumor dataset [207], [208]:** The dataset provides a collection of 8798 MRI images from 20 different patients with primary newly diagnosed glioblastoma. The images are available in DICOM format and contain T1w (pre and post-contrast agent), FLAIR, T2w, ADC, normalized cerebral blood flow, normalized relative cerebral blood volume, standardized relative cerebral blood volume, and binary tumor masks (generated using T1w images).

3.8 Gastric Cancer Dataset

Some publicly available datasets for gastric cancer detection include:

- **AIDA-E [209]:** The dataset provides images from three different categories including (i) normal mucosal regions, which is termed as normal class,

and a collection of abnormal images with (i) regions belonging to group II (metaplasia), and (iii) group III (dysplasia). The normal images class is composed of 56 images while the abnormal class contains a total of 120 images, where 96 images belongs to group II (i.e., metaplasia) and the remaining 24 images belong to group III (i.e., dysplasia).

- **Gastric Cancer Dataset [210]:** The data is composed of 926 images from 58 different patients. All the images are extracted from a video each of size 1000×870 . A total of 228 images are abnormal (i.e., with cancer) while the normal set is composed of 698 images. All the images are manually analyzed and annotated by experts.
- **Gastric Cancer Dataset [211]:** The dataset provides a total of 11,686 endoscopic images. The images are divided into three categories, namely T1a-EGC, T1a-EGC, and non-EGC containing 1097, 1005, and 9834 images, respectively. The third category (i.e., non-EGC) is endoscopic images of the gastric mucosa containing chronic gastritis, chronic atrophic gastritis, intestinal metaplasia, and erosion. All the images are manually analyzed and annotated by the domain experts.

3.9 PKU/MCADD Dataset

No public dataset found...

-
-
-

3.10 Heart Disease Dataset

Some publicly available dataset in the domain include:

- **Heart Disease Dataset [212]:** The dataset provides a large collection of heart patients data. In total, the dataset covers 76 attributes. All the data samples are annotated either 0 representing the absence of heart disease or 1 indicating the presence of heart diseases in the patient.
- **Heart Disease Dataset [213]:** The dataset provides a collection of medical records of 299 heart patients including 105 female and 194 male patients in the range of 40 and 95 years old. Each data sample is composed of 12 features including *gender*, *smoking*, *diabetes*, *blood pressure*, *anaemia*, *age*, *ejection fraction*, *sodium*, *creatinine*, *platelets*, and *CPK*.
- **Kaggle Heart Disease Dataset [214]:** The dataset is composed of 4,000 medical records with 15 different attributes. The attributes can be divided into five different groups, namely (i) demographic, (ii) education, (iii) behavioral, (iv) medical history, and (v) current medical conditions. The medical history group is composed of blood pressure medicine,

prevalent stroke, prevalent hyp, and diabetes. The current medical condition group contains total cholesterol level, systolic blood pressure, heart rate, and blood glucose level.

- **Heart statlog [215]:** This dataset is composed of a large collection of medical records covering 13 different attributes. These attributes include age, sex, chest pain type, resting blood pressure, serum cholesterol, fasting blood sugar, resting electrocardiographic results, maximum heart rate achieved, exercise-induced angina, depression induced by exercise relative to rest, the slope of the peak exercise ST segment, and the number of major vessels.

3.11 Diabetes Dataset

Some publicly available dataset in of diabetes detection include:

- **Diabetes Dataset [216]:** The dataset provides a large collection of medical records obtained from two different sources including automatic records from an electronic device as well as data taken from paper records. The timing of meals is crucial for diabetes readings. To this aim, the electronic device provides exact timestamp events through a clock while the paper records provide logical time slots, such as breakfast, lunch, dinner, and bedtime, where specific timings are set for each of these activities. Each instance of the records provides four different types of readings, namely date, time, code, and value/readings. The code can have 16 different possible values relevant to diabetes.
- **Pima Indians Diabetes Database [217]:** This is one of the mostly widely used diabetes dataset for training and evaluation ML techniques. The dataset provides a large collection of medical records collected at National Institute of Diabetes and Digestive and Kidney Diseases India. The key features/readings provided by the dataset include *pregnancies (for female only)*, *glucose*, *blood pressure*, *skin thickness*, *insulin in take*, *BMI*, *diabetes pedigree function*, and *age*. The dataset mainly provides two classes indicating the presence and absence of diabetes in a patient.

3.12 Alzheimer's Disease Dataset

Some publicly available dataset in the domain include:

- **Alzheimer's Disease Neuroimaging Initiative (ADNI) Data [218]:** ADNI provides a platform for the researchers to work jointly by collaborating in data collection and validation. It provides a valuable resource for relevant data in the form of MRI and Positron emission tomography (PET) images, genetics, cognitive tests, Cerebrospinal fluid (CSF), and blood biomarkers.

- **Alzheimer's Disease and Healthy Aging Data Portal [219]:** The portal provides an access to national and state-level data on several health and well-being issues including Alzheimer's disease. The portal allows crawling Alzheimer's disease and healthy aging data by location and indicators. These indicators include caregiving, subjective cognitive decline, screenings and vaccinations, mental health, nutrition, and physical activity, smoking and alcohol use, and overall health.
- **OASIS-1 [220]:** The dataset is developed as part of the Open Access Series of Imaging Studies (OASIS) project, and contains a large collection of MRI images. The population of the patients/subjects include both men and female with right hand dominance from each category. A total of 373 scanning sessions were conducted by scanning each subject at least two times with a gap of one year. The feature set of the dataset is composed of 10 features, namely *age, gender, MRI delay, socioeconomic status, mini mental state examination score, clinical dementia rating, atlas scaling factor, estimated total intracranial volume, and normalized whole brain volume*.
- **NHIS-NSC [221]:** The dataset provides medical records of around 1,025,340 subjects. The dataset covers a wide range of features including readings related to individual services, diagnoses, prescriptions, demographics, income levels, disease and medication codes, health profiles, and medical history.

3.13 HIV Dataset

Some publicly available dataset in the domain include:

- **HIV-1 protease cleavage Data Set [222]:** The dataset provides a large collection of medical records, which are available in four different versions. The key features covered by the dataset include *amino acid, G (Glycine), P (Proline), A (Alanine), V (Valine), L (Leucine), I (Isoleucine), M (Methionine), C (Cysteine), F (Phenylalanine), Y (Tyrosine), W (Tryptophan), H (Histidine), K (Lysine), R (Arginine), Q (Glutamine), N (Asparagine), E (Glutamic Acid), D (Aspartic Acid), S (Serine), and T (Threonine)*. All the samples are either annotated as +1 (i.e., cleaved) or -1 not cleaved.
- **HIV Dataset [223]:** The dataset is based on the Demographic and Health Surveys (DHS) surveys conducted in 2013 in ten East and Southern African countries with a generalized HIV epidemic, namely Angola, Burundi, Ethiopia, Lesotho, Malawi, Mozambique, Namibia, Rwanda, Zambia and, Zimbabwe. The dataset covers data from both male and female subjects having 68,979 and 83,910 samples, respectively.

3.14 HPV Dataset

Some publicly available dataset in the domain include:

- **Cervical cancer (Risk Factors) Dataset [224]:** The dataset provides medical records of 858 patients in terms of demographic information, habits, and historic medical records. In total, the feature set is composed of 36 different types of readings. It is important to mention that the dataset has several missing values for different factors mainly due to the refusal of the patient to answer certain questions on the basis of privacy risks. Such missing values are replaced with sample mean values, which is one of the main limitations of the dataset.

3.15 Obesity Dataset

Some publicly available dataset in the domain include:

- **RISKESDAS [225]:** The dataset is based on a survey conducted by the Indonesian Ministry of Health. The survey is regularly conducted by a gap of five years across 34 provinces and 514 districts/cities to highlight and keep track of public health indicators. For the collection of an obesity dataset, Thamrin et al. [225] collected a subset of the samples by removing the missing values, such as BMI. Moreover, 21 different features having a direct or indirect impact on obesity are chosen.
- **Obesity Dataset [226]:** The dataset covers data from three different countries, namely Mexico, Peru, and Colombia based on the eating habits and physical conditions of the residents. In total, the dataset is composed of 2111 samples each covering 17 different features. The dataset is annotated using seven different labels namely (i) insufficient weight, (ii) normal weight, (iii) overweight level I, (iv) overweight level II, (v) obesity type I, (vi) obesity type II, and (vii) obesity type III.
- **Obesity Dataset [227]:** The dataset is based on a survey conducted on students using 18 different factors related to obesity. A total of 178 students, including 81 male and 97 female, participated in the survey. All the participants of the survey belong to the 18 to 25 years old group.

3.16 Lung Cancer Dataset

Some publicly available dataset in the domain include:

- **Lung Cancer PET/CT Dataset [228]:** The dataset provides a large collection of CT and PET-CT DICOM images of lung cancer patients. All the images are analyzed and pre-processed on the mediastinum (window width, 350 HU; level, 40 HU) and lung (window width, 1,400 HU; level, -700 HU) settings. Moreover, the images were re-constructed in 2mm-slice-thick and lung settings. The annotation was performed by domain experts.

- **ELCAP Public Lung Image Database:** The database is developed as part of a collaboration between the ELCAP and VIA research groups. It contains a total of 50 low-dose documented whole-lung CT scans for detection, which are obtained with a 1.25 mm slice thickness. The CT scans are annotated by domain experts.
- **NELSON Dataset [229]:** The dataset provides separate collections of CT scans and alphanumeric data describing the participant's characteristics, test results along with the CT scan annotations. The CT scans are captured with four different systems including three Sensation-16, Siemens Medical Solutions, Forchheim, Germany M8000 IDT CT systems and one Brilliance 16P, Philips Medical Systems, Cleveland. The CT scans are analyzed and annotated by domain experts.
-

4 LESSONS LEARNED, CHALLENGES AND FUTURE RESEARCH DIRECTIONS

4.1 Lessons Learned

4.2 Challenges and Future Research Directions

- **Interpretable AI Solutions:** In conventional AI-based healthcare solutions, AI is used as a black box. However, to make AI-based solutions more trust-able and accountable, interpretation and explanations of AI predictions are crucial. The deployment of interpretable AI solutions will not only enhance transparency and fairness but will also help in building stakeholders' trust. The literature already reports some efforts in this direction [230].
- **Security and Privacy:** ML algorithms are subject to several types of security attacks, such as adversarial attacks, model extraction, Trojan, membership inference, and model inversion attacks [231], [232]. Healthcare is one of the key applications where a systematic approach for the security and the development of adversarial robust AI algorithms is crucial. Moreover, patient privacy is another key aspect of smart healthcare. To this aim, the literature already reports some interesting efforts [233]. In this regard, Federated Learning (FL), which allows training an ML model collaboratively without sharing data with each other, is one of the promising solutions.
- **Ethical Issues:** An important issue pertaining to sensors is ethics. More options are available when implanted sensors are available for consideration, but more ethical concerns arise regarding obtaining information and patient consent [234]. There are also ethical concerns regarding healthcare decisions made from algorithms and the level of decision-making from providers. The question

must be answered [235], *When Should AI recommendations be used and when should they be rejected?*

- **Data Auditing:** Data auditing is one of the most crucial phases of the development and deployment of AI solutions in human-centric applications in general and healthcare in particular [231]. In smart health applications, data is generally collected from different sources and devices. Thus, it is recommended to pay proper attention to analyzing the risk associated with the data and its potential impact on the model.
- **Interoperability and Integration:** One of the key challenges and barriers in adopting the insights obtained from the research and the resulted new techniques in practice is the lack of data infrastructure [236]. The newly proposed AI-based solutions generally work and are evaluated on carefully curated sets of health data. However, in practice, the data is heterogeneous, messy, noisy, fragmented, and hard to access.
- **An access to Centralized Data:**
to be added
- **Understanding Disease Causality:** One of the main goals of epidemiology and biomedical applications is to identify causal relationships between outcomes - like death, diseases, or injuries - and exposures - like smoking cigarettes, eating junk food, or drinking alcohol [237]. Most machine learning-based methods have been proved to be efficient to predict outcomes rather than determining the causality [238]. Identifying co-causal factors in such disease processes and eventually understanding their different mechanisms of action will be challenging. Efforts to address these problems mostly include training machine learning models on more examples. But as the environment grows in complexity, it becomes impossible to cover the entire distribution by adding more training examples. During the pandemics such as COVID-19, many machine learning systems usually fail because they have been trained on statistical regularities instead of causal relations [239]. As life patterns changed, the accuracy of the models dropped. To overcome these challenges, causal machine learning offers a promising solutions [240] to improving the accuracy of medical diagnosis and address machine learning's lack of generalization when interventions change the statistical distributions of a problem. It allows us to respond to situations we have not seen before and think about counterfactuals.
-
-

5 DISCUSSION

TO BE EDITED Machine Learning is an emerging topic of computer science that is quickly integrating into many areas of professionalism, the medical field included. As a result, many corners of the medical profession, including detecting and identifying disease and risk factors, are having ML principles be applied to them to automate processes typically done by humans, resulting in quicker results and time saved for both the patient and the medical practitioner. The goal of these implementations is to benefit all parties involved in these medical situations with fast and reliable identifications and diagnoses to begin early patient intervention and allow professionals to spend additional time with more patients.

The information contained in this paper shows that classification-based ML and DL models seem to be ideal for predicting, detecting, and classifying diseases. A RF obtained an accuracy of 99% for detecting epileptic patients from healthy controls using EEG data. A RF and LR obtained an accuracy of 99.83% for detecting chronic kidney disease. A SVM detected Autism with 95% accuracy. A RF classified anxiety with 100% accuracy; however, the dataset was small and unbalanced, and a RBFN classified anxiety with 97.48% accuracy. A DTL-MC produced the highest model accuracy of 92.64% for the detection of Covid-19 cough detection. MRSA risk and early diagnosis was best detected using binary classification achieving an $AUC < .95$ for detection days before diagnosis and an AUC of .98 for detection one day prior to diagnosis. Brain tumors were accurately classified 96.3% of the time using a KNN classifier on MRI data.

Gastric cancer was accurately detected using an SVM classifier with 97.13% accuracy from saliva samples using Raman spectra data for biomarker classification. PKU and MCADD were accurately classified using LR and achieved a sensitivity $> 96.8\%$ and specificity $> 99.98\%$. Cardiovascular risk was detected using ML models on the NHANES dataset and the Framingham Heart Study and showed the best accuracy of 98.5% in 214 seconds from the RF model. ANNs had the highest accuracy (85%) for cardiac arrest prediction. For diabetes detection LR performed best on the PIMA Indian Diabetes Dataset with 77.9% accuracy. A stacked LSTM model achieved an accuracy of 97.5% in detecting Parkinson's disease based on gait data. DL and SVM achieved an accuracy of 87% in classifying Alzheimer's Disease using MRI images, EHRs, and SNPs. HIV classification achieved an accuracy of 97% using the Kameris ML software package with the highest performance coming from the SVM, MLP, and LR algorithms. HPV association in patients with OPSCC achieved a prediction score with an AUC of 0.74. Obesity level classification was determined by an ensemble model composed of a GLM, RF, and partial least squares, achieving an accuracy of 89.68%. Lung cancer prediction achieved an accuracy of 96.67% using an ANN model applied on text-based data.

Future work can consider adding more neurological diseases to distinguish from epilepsy detection rather than just classifying epilepsy from healthy controls. Future work for Chronic Kidney disease can consider using image-based data for detection. For anxiety, future work should consider using larger datasets with more diverse or different questions for comparison. COVID-19 research is critical to understand with the event of its recent introduction into the healthcare field, causing a global pandemic. The availability of data pertaining to this disease is new and limited, but the occurrence of future cases will provide more information, leading to new developments in the accurate detection of the disease, as well as identifying risk factors leading to mortality which can ultimately improve treatment.

For MRSA detection and prediction, future work may consider using multiple hospitals rather than just one that was used in the present study to create a more generalizable, all-encompassing prediction. For brain cancer, future research should consider using a larger dataset from multiple sources for more generalizability. Future work regarding PKU and MCADD classification should consider more up-to-date research, since the study discussed in this paper was performed in 2004.

The Framingham Heart Study is a multigenerational, longitudinal study that can provide new insights for heart disease research. Using datasets such as the Framingham study in detecting and evaluating factors pertaining to cardiovascular disease provides a rigorous design incorporating genetic, behavioral, and environmental factors to provide exemplary analysis and should be evaluated periodically so that any changes can be accurately integrated. The PIMA Indian diabetes dataset was used to evaluate different ML techniques for diabetes detection, providing an evaluation that could be extended for larger, more diverse datasets. The PIMA dataset only observes female ages 21 and older with PIMA Indian heritage, and future research should incorporate a more diverse population for a more comprehensive analysis.

Future work regarding gait abnormality detection in older adults should consider spatial parameters in addition to plantar pressure measurement and inertial measurement units. Future research should also consider addressing issues pertaining to missing data, overestimation, demographic variability, and larger data amounts. More research on DDRI involving the application of various algorithms (and combinations of algorithms), feature selection techniques, etc. will ideally produce optimal models. Future research regarding Alzheimer's detection should include testing on a larger sample size. HIV classifiers were not able to have performance explanations and acted as a black-box, so future work may attempt to provide explanations for the classifications. Future research regarding HPV classification can focus on resolving the interrater reliability issue. Speed of obesity level classification for ML models can be improved with further study

on the subject. Future research regarding lung cancer prediction using text-based data could incorporate a wider range of data.

Future research should consider focusing on improving and developing techniques such as combining multiple ML/DL methods to enhance audio-based and image-based classification problems and thus normalizing the use of these practices in the medical field. Currently, the use of ML models and biomedical sensors is considered a new process and is still gaining acceptance, making the implementation of these practices questionable to some practitioners. Future studies incorporating the use of innovative research pertaining to the improvement ML techniques will inevitably lead to universal acceptance of ML applications in the healthcare field.

6 CONCLUSIONS

REFERENCES

- [1] N. Alexi, J. Hvam, B. W. Lund, L. Nsubuga, R. M. de Oliveira Hansen, K. Thamsborg, F. Lofink, D. V. Byrne, and J. J. Leisner, "Potential of novel cadaverine biosensor technology to predict shelf life of chilled yellowfin tuna (*thunnus albacares*)," *Food Control*, vol. 119, p. 107458, 2020.
- [2] B. M. Bates-Jensen, H. E. McCreath, and A. Patlan, "Subepidermal moisture detection of pressure induced tissue damage on the trunk: the pressure ulcer detection study outcomes," *Wound Repair and Regeneration*, vol. 25, no. 3, pp. 502–511, 2017.
- [3] M. Clinic, "Diabetes," 2020.
- [4] A. Tahamtan and A. Ardebili, "Real-time rt-pcr in covid-19 detection: issues affecting the results," 2020.
- [5] S. Shomaji, P. Dehghanzadeh, A. Roman, D. Forte, S. Bhunia, and S. Mandal, "Early detection of cardiovascular diseases using wearable ultrasound device," *IEEE Consumer Electronics Magazine*, vol. 8, no. 6, pp. 12–21, 2019.
- [6] B. Bohunicky and S. A. Mousa, "Biosensors: the new wave in cancer diagnosis," *Nanotechnology, science and applications*, vol. 4, p. 1, 2011.
- [7] A. T. Hirsch, M. H. Criqui, D. Treat-Jacobson, J. G. Regensteiner, M. A. Creager, J. W. Olin, S. H. Krook, D. B. Hunninghake, A. J. Comerota, M. E. Walsh, et al., "Peripheral arterial disease detection, awareness, and treatment in primary care," *Jama*, vol. 286, no. 11, pp. 1317–1324, 2001.
- [8] R. J. Genco and W. S. Borgnakke, "Risk factors for periodontal disease," *Periodontology 2000*, vol. 62, no. 1, pp. 59–94, 2013.
- [9] N. Bhalla, P. Jolly, N. Formisano, and P. Estrela, "Introduction to biosensors," *Essays Biochem*, vol. 60, pp. 1–8, 06 2016.
- [10] P. Mehrotra, "Biosensors and their applications—a review," *Journal of oral biology and craniofacial research*, vol. 6, no. 2, pp. 153–159, 2016.
- [11] J. Kim, A. S. Campbell, B. E.-F. de Ávila, and J. Wang, "Wearable biosensors for healthcare monitoring," *Nature biotechnology*, vol. 37, no. 4, pp. 389–406, 2019.
- [12] A. Villegas, D. McEneaney, and O. Escalona, "Arm-ecg wireless sensor system for wearable long-term surveillance of heart arrhythmias," *Electronics*, vol. 8, no. 11, p. 1300, 2019.
- [13] N. S. Rajliwall, R. Davey, and G. Chetty, "Machine learning based models for cardiovascular risk prediction," in *2018 International Conference on Machine Learning and Data Engineering (iCMLDE)*, pp. 142–148, IEEE, 2018.
- [14] J. Shen, C. J. Zhang, B. Jiang, J. Chen, J. Song, Z. Liu, Z. He, S. Y. Wong, P.-H. Fang, and W.-K. Ming, "Artificial intelligence versus clinicians in disease diagnosis: Systematic review," *JMIR medical informatics*, vol. 7, no. 3, p. e10010, 2019.
- [15] J. Kerns, "Are wearables decentralizing medicine?," 2019.
- [16] M. A. Myszczyńska, P. N. Ojamies, A. M. Lacoste, D. Neil, A. Saffari, R. Mead, G. M. Hautbergue, J. D. Holbrook, and L. Ferraiuolo, "Applications of machine learning to diagnosis and treatment of neurodegenerative diseases," *Nature Reviews Neurology*, vol. 16, no. 8, pp. 440–456, 2020.
- [17] M. Bracher-Smith, K. Crawford, and V. Escott-Price, "Machine learning for genetic prediction of psychiatric disorders: a systematic review," *Molecular Psychiatry*, pp. 1–10, 2020.
- [18] R. Alizadehsani, A. Khosravi, M. Roshanzamir, M. Abdar, N. Sarrafzadegan, D. Shafie, F. Khozeimeh, A. Shoeibi, S. Nahavandi, M. Panahi-azar, et al., "Coronary artery disease detection using artificial intelligence techniques: A survey of trends, geographical differences and diagnostic features 1991-2020," *Computers in Biology and Medicine*, p. 104095, 2020.
- [19] G. Saranya and A. Pravin, "A comprehensive study on disease risk predictions in machine learning," *International Journal of Electrical and Computer Engineering*, vol. 10, no. 4, p. 4217, 2020.
- [20] Y. Khan, U. Qamar, N. Yousaf, and A. Khan, "Machine learning techniques for heart disease datasets: a survey," in *Proceedings of the 2019 11th International Conference on Machine Learning and Computing*, pp. 27–35, 2019.
- [21] A. Choudhury and D. Gupta, "A survey on medical diagnosis of diabetes using machine learning techniques," in *Recent Developments in Machine Learning and Data Analytics*, pp. 67–78, Springer, 2019.
- [22] A. B. Shatte, D. M. Hutchinson, and S. J. Teague, "Machine learning in mental health: a scoping re-

- view of methods and applications," *Psychological medicine*, vol. 49, no. 9, pp. 1426–1448, 2019.
- [23] S. T. H. Kieu, A. Bade, M. H. A. Hijazi, and H. Kollivand, "A survey of deep learning for lung disease detection on medical images: State-of-the-art, taxonomy, issues and future directions," *Journal of Imaging*, vol. 6, no. 12, p. 131, 2020.
- [24] N. C. Caballé, J. L. Castillo-Sequera, J. A. Gómez-Pulido, J. M. Gómez-Pulido, and M. L. Polo-Luque, "Machine learning applied to diagnosis of human diseases: A systematic review," *Applied Sciences*, vol. 10, no. 15, p. 5135, 2020.
- [25] S. Panicker *et al.*, "Use of machine learning techniques in healthcare: A brief review of cardiovascular disease classification," 2020.
- [26] M. A. Ebrahimighahnavieh, S. Luo, and R. Chiong, "Deep learning to detect alzheimer's disease from neuroimaging: A systematic literature review," *Computer methods and programs in biomedicine*, vol. 187, p. 105242, 2020.
- [27] A. Channa, N. Popescu, and V. Ciobanu, "Wearable solutions for patients with parkinson's disease and neurocognitive disorder: A systematic review," *Sensors*, vol. 20, no. 9, p. 2713, 2020.
- [28] K. K. Hyde, M. N. Novack, N. LaHaye, C. Parlett-Pelleriti, R. Anden, D. R. Dixon, and E. Linstead, "Applications of supervised machine learning in autism spectrum disorder research: a review," *Review Journal of Autism and Developmental Disorders*, vol. 6, no. 2, pp. 128–146, 2019.
- [29] S. Uddin, A. Khan, M. E. Hossain, and M. A. Moni, "Comparing different supervised machine learning algorithms for disease prediction," *BMC Medical Informatics and Decision Making*, vol. 19, no. 1, pp. 1–16, 2019.
- [30] T. Jo, K. Nho, and A. J. Saykin, "Deep learning in alzheimer's disease: diagnostic classification and prognostic prediction using neuroimaging data," *Frontiers in aging neuroscience*, vol. 11, p. 220, 2019.
- [31] S. Safdar, S. Zafar, N. Zafar, and N. F. Khan, "Machine learning based decision support systems (dss) for heart disease diagnosis: a review," *Artificial Intelligence Review*, vol. 50, no. 4, pp. 597–623, 2018.
- [32] Z. Xiong, T. Liu, G. Tse, M. Gong, P. A. Gladding, B. H. Smaill, M. K. Stiles, A. M. Gillis, and J. Zhao, "A machine learning aided systematic review and meta-analysis of the relative risk of atrial fibrillation in patients with diabetes mellitus," *Frontiers in physiology*, vol. 9, p. 835, 2018.
- [33] E. Pellegrini, L. Ballerini, M. d. C. V. Hernandez, F. M. Chappell, V. González-Castro, D. Anblagan, S. Danso, S. Muñoz-Maniega, D. Job, C. Pernet, *et al.*, "Machine learning of neuroimaging for assisted diagnosis of cognitive impairment and dementia: a systematic review," *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, vol. 10, pp. 519–535, 2018.
- [34] M. Bakator and D. Radosav, "Deep learning and medical diagnosis: A review of literature," *Multimodal Technologies and Interaction*, vol. 2, no. 3, p. 47, 2018.
- [35] E. Cuadrado-Godia, P. Dwivedi, S. Sharma, A. O. Santiago, J. R. Gonzalez, M. Balcells, J. Laird, M. Turk, H. S. Suri, A. Nicolaides, *et al.*, "Cerebral small vessel disease: a review focusing on pathophysiology, biomarkers, and machine learning strategies," *Journal of stroke*, vol. 20, no. 3, p. 302, 2018.
- [36] K. Shailaja, B. Seetharamulu, and M. Jabbar, "Machine learning in healthcare: A review," in *2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, pp. 910–914, IEEE, 2018.
- [37] J. Fagerström, M. Bång, D. Wilhelms, and M. S. Chew, "Lisep lstm: a machine learning algorithm for early detection of septic shock," *Scientific reports*, vol. 9, no. 1, pp. 1–8, 2019.
- [38] F. Khoshnevisan, J. Ivy, M. Capan, R. Arnold, J. Huddleston, and M. Chi, "Recent temporal pattern mining for septic shock early prediction," in *2018 IEEE International Conference on Healthcare Informatics (ICHI)*, pp. 229–240, 2018.
- [39] D. Spathis and P. Vlamos, "Diagnosing asthma and chronic obstructive pulmonary disease with machine learning," *Health informatics journal*, vol. 25, no. 3, pp. 811–827, 2019.
- [40] R. Buettner, J. Frick, and T. Rieg, "High-performance detection of epilepsy in seizure-free eeg recordings: A novel machine learning approach using very specific epileptic eeg sub-bands," in *ICIS*, 2019.
- [41] E. Molina, C. E. S. Torres, R. Salazar-Cabrera, D. M. López, and R. Vargas-Cañas, "Intelligent telehealth system to support epilepsy diagnosis," *Journal of Multidisciplinary Healthcare*, vol. 13, p. 433, 2020.
- [42] G. Sahebzamani, M. Saffar, and H. Soltanian-Zadeh, "Machine learning based analysis of structural mri for epilepsy diagnosis," in *2019 4th International Conference on Pattern Recognition and Image Analysis (IPRIA)*, pp. 58–63, 2019.
- [43] A. Sobrinho, A. C. M. D. S. Queiroz, L. Dias Da Silva, E. De Barros Costa, M. Eliete Pinheiro, and A. Perkusich, "Computer-aided diagnosis of chronic kidney disease in developing countries: A comparative analysis of machine learning techniques," *IEEE Access*, vol. 8, pp. 25407–25419, 2020.
- [44] J. Qin, L. Chen, Y. Liu, C. Liu, C. Feng, and B. Chen, "A machine learning methodology for diagnosing chronic kidney disease," *IEEE Access*, vol. 8, pp. 20991–21002, 2020.
- [45] R. A. Allassaf, K. A. Alsulaim, N. Y. Alroomi, N. S. Alsharif, M. F. Aljubeir, S. O. Olatunji, A. Y. Alahmadi, M. Imran, R. A. Alzahrani, and N. S. Al-

- turayef, "Preemptive diagnosis of chronic kidney disease using machine learning techniques," in *2018 International Conference on Innovations in Information Technology (IIT)*, pp. 99–104, 2018.
- [46] H. Polat, H. D. Mehr, and A. Cetin, "Diagnosis of chronic kidney disease based on support vector machine by feature selection methods," *Journal of medical systems*, vol. 41, no. 4, p. 55, 2017.
- [47] Y. Zhang, F. Li, F. Yuan, K. Zhang, L. Huo, Z. Dong, Y. Lang, Y. Zhang, M. Wang, Z. Gao, *et al.*, "Diagnosing chronic atrophic gastritis by gastroscopy using artificial intelligence," *Digestive and Liver Disease*, vol. 52, no. 5, pp. 566–572, 2020.
- [48] S. Syed, S. Sengupta, L. Ehsan, E. Bonkowski, C. Moskaluk, A. Griffiths, A. Otle, A. Patel, B. Kirschner, D. Mack, *et al.*, "Machine learning for crohn's disease phenotype modeling using biopsy images," *Inflammatory Bowel Diseases*, vol. 27, no. Supplement_1, pp. S10–S11, 2021.
- [49] H. M. Khorasani, H. Usefi, and L. Peña-Castillo, "Detecting ulcerative colitis from colon samples using efficient feature selection and machine learning," *Scientific Reports*, vol. 10, no. 1, pp. 1–9, 2020.
- [50] H. Li, L. Lai, and J. Shen, "Development of a susceptibility gene based novel predictive model for the diagnosis of ulcerative colitis using random forest and artificial neural network," *Aging (Albany NY)*, vol. 12, no. 20, p. 20471, 2020.
- [51] F. K. Alsheref and W. H. Gomaa, "Blood diseases detection using classical machine learning algorithms," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 10, no. 7, 2019.
- [52] C.-Y. Chen, W.-C. Lin, and H.-Y. Yang, "Diagnosis of ventilator-associated pneumonia using electronic nose sensor array signals: solutions to improve the application of machine learning in respiratory research," *Respiratory research*, vol. 21, no. 1, p. 45, 2020.
- [53] C.-H. Hsieh, R.-H. Lu, N.-H. Lee, W.-T. Chiu, M.-H. Hsu, and Y.-C. J. Li, "Novel solutions for an old disease: diagnosis of acute appendicitis with random forest, support vector machines, and artificial neural networks," *Surgery*, vol. 149, no. 1, pp. 87–93, 2011.
- [54] E. Aydin, İ. U. Türkmen, G. Namli, Ç. Öztürk, A. B. Esen, Y. N. Eray, E. Eroğlu, and F. Akova, "A novel and simple machine learning algorithm for preoperative diagnosis of acute appendicitis in children," *Pediatric surgery international*, vol. 36, no. 6, pp. 735–742, 2020.
- [55] O. F. Akmes, G. Dogan, H. Kor, H. Erbay, and E. Demir, "The use of machine learning approaches for the diagnosis of acute appendicitis," *Emergency medicine international*, vol. 2020, 2020.
- [56] I. A. Ozkan, M. Koklu, and I. U. Sert, "Diagnosis of urinary tract infection based on artificial intelligence methods," *Computer methods and programs in biomedicine*, vol. 166, pp. 51–59, 2018.
- [57] J. Oh, B.-L. Oh, K.-U. Lee, J.-H. Chae, and K. Yun, "Identifying schizophrenia using structural mri with a deep learning algorithm," *Frontiers in psychiatry*, vol. 11, p. 16, 2020.
- [58] L.-A. Claude, J. Houenou, E. Duchesnay, and P. Favre, "Will machine learning applied to neuroimaging in bipolar disorder help the clinician? a critical review and methodological suggestions," *Bipolar disorders*, vol. 22, no. 4, pp. 334–355, 2020.
- [59] K.-S. Na, S.-E. Cho, and S.-J. Cho, "Machine learning-based discrimination of panic disorder from other anxiety disorders," *Journal of Affective Disorders*, vol. 278, pp. 1–4, 2021.
- [60] B. Mwangi, K. P. Ebmeier, K. Matthews, and J. Douglas Steele, "Multi-centre diagnostic classification of individual structural neuroimaging scans from patients with major depressive disorder," *Brain*, vol. 135, no. 5, pp. 1508–1521, 2012.
- [61] P. Kumar, S. Garg, and A. Garg, "Assessment of anxiety, depression and stress using machine learning models," *Procedia Computer Science*, vol. 171, pp. 1989–1998, 2020.
- [62] S. Wshah, C. Skalka, and M. Price, "Predicting posttraumatic stress disorder risk: a machine learning approach," *JMIR mental health*, vol. 6, no. 7, p. e13946, 2019.
- [63] G. Tartarisco, G. Cicceri, D. Di Pietro, E. Leonardi, S. Aiello, F. Marino, F. Chiarotti, A. Gagliano, G. M. Arduino, F. Apicella, *et al.*, "Use of machine learning to investigate the quantitative checklist for autism in toddlers (q-chat) towards early autism screening," *Diagnostics*, vol. 11, no. 3, p. 574, 2021.
- [64] C. A. Stamatis, M. C. Batistuzzo, T. Tanamatis, E. C. Miguel, M. Q. Hoexter, and K. R. Timpano, "Using supervised machine learning on neuropsychological data to distinguish ocd patients with and without sensory phenomena from healthy controls," *British Journal of Clinical Psychology*, vol. 60, no. 1, pp. 77–98, 2021.
- [65] H. Christiansen, M.-L. Chavanon, O. Hirsch, M. H. Schmidt, C. Meyer, A. Müller, H.-J. Rumpf, I. Grigorev, and A. Hoffmann, "Use of machine learning to classify adult adhd and other conditions based on the conners' adult adhd rating scales," *Scientific reports*, vol. 10, no. 1, pp. 1–10, 2020.
- [66] M. Barstugan, U. Ozkaya, and S. Ozturk, "Coronavirus (covid-19) classification using ct images by machine learning methods," *arXiv preprint arXiv:2003.09424*, 2020.
- [67] P. K. Sethy, S. K. Behera, P. K. Ratha, and P. Biswas, "Detection of coronavirus disease (covid-19) based on deep features and support vector machine," 2020.
- [68] M. Toğaçar, B. Ergen, Z. Cömert, and F. Özyurt, "A

- deep feature learning model for pneumonia detection applying a combination of mrmr feature selection and machine learning models," *Irbm*, vol. 41, no. 4, pp. 212–222, 2020.
- [69] V. M. Lelis, E. Guzman, and M.-V. Belmonte, "Non-invasive meningitis diagnosis using decision trees," *IEEE Access*, vol. 8, pp. 18394–18407, 2020.
- [70] S. O. Alile and M. E. Bello, "A machine learning approach for diagnosing meningococcal meningitis," *Int. J. Sci. Res. in Computer Science and Engineering Vol*, vol. 8, no. 3, 2020.
- [71] A. Yahiaoui, O. Er, and N. Yumusak, "A new method of automatic recognition for tuberculosis disease diagnosis using support vector machines," *Biomedical Research (0970-938X)*, vol. 28, no. 9, 2017.
- [72] S. Khan, R. Ullah, S. Shahzad, N. Anbreen, M. Bilal, and A. Khan, "Analysis of tuberculosis disease through raman spectroscopy and machine learning," *Photodiagnosis and photodynamic therapy*, vol. 24, pp. 286–291, 2018.
- [73] S. S. Meraj, R. Yaakob, A. Azman, S. N. M. Rum, and A. Nazri, "Artificial intelligence in diagnosing tuberculosis: A review," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 9, no. 1, pp. 81–91, 2019.
- [74] B. N. Narayanan, R. Ali, and R. C. Hardie, "Performance analysis of machine learning and deep learning architectures for malaria detection on cell images," in *Applications of Machine Learning*, vol. 11139, p. 111390W, International Society for Optics and Photonics, 2019.
- [75] S. Gambhir, S. K. Malik, and Y. Kumar, "The diagnosis of dengue disease: An evaluation of three machine learning approaches," *International Journal of Healthcare Information Systems and Informatics (IJHISI)*, vol. 13, no. 3, pp. 1–19, 2018.
- [76] A. E. Tozzi, F. Gesualdo, C. Rizzo, E. Carloni, L. Russo, I. Campagna, A. Villani, A. Reale, C. Concato, G. Linardos, *et al.*, "A data driven clinical algorithm for differential diagnosis of pertussis and other respiratory infections in infants," *Plos one*, vol. 15, no. 7, p. e0236041, 2020.
- [77] M. L. M. De Souza, G. A. Lopes, A. C. Branco, J. K. Fairley, and L. A. D. O. Fraga, "Leprosy screening based on artificial intelligence: Development of a cross-platform app," *JMIR mHealth and uHealth*, vol. 9, no. 4, p. e23718, 2021.
- [78] Y. Xu, C. Li, Y. Jiang, M. Guo, Y. Yang, Y. Yang, and H. Yu, "Electrochemical impedance spectroscopic detection of e. coli with machine learning," *Journal of The Electrochemical Society*, vol. 167, no. 4, p. 047508, 2020.
- [79] M. Fatima, M. Pasha, *et al.*, "Survey of machine learning algorithms for disease diagnostic," *Journal of Intelligent Learning Systems and Applications*, vol. 9, no. 01, p. 1, 2017.
- [80] J. S. Sartakhti, M. H. Zangoeei, and K. Mozafari, "Hepatitis disease diagnosis using a novel hybrid method based on support vector machine and simulated annealing (svm-sa)," *Computer methods and programs in biomedicine*, vol. 108, no. 2, pp. 570–579, 2012.
- [81] C. F. O. R. Melo, L. C. Navarro, D. N. De Oliveira, T. M. Guerreiro, E. d. O. Lima, J. Delafiori, M. Z. Dabaja, M. d. S. Ribeiro, M. de Menezes, R. G. M. Rodrigues, *et al.*, "A machine learning application based in random forest for integrating mass spectrometry-based metabolomic data: a simple screening method for patients with zika virus," *Frontiers in bioengineering and biotechnology*, vol. 6, p. 31, 2018.
- [82] E. Marquez and V. Barrón, "Artificial intelligence system to support the clinical decision for influenza," in *2019 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC)*, pp. 1–5, 2019.
- [83] T. Hartvigsen, C. Sen, S. Brownell, E. Teeple, X. Kong, and E. A. Rundensteiner, "Early prediction of mrsa infections using electronic health records," in *HEALTHINF*, pp. 156–167, 2018.
- [84] E. I. Zacharaki, V. G. Kanas, and C. Davatzikos, "Investigating machine learning techniques for mri-based classification of brain neoplasms," *International journal of computer assisted radiology and surgery*, vol. 6, no. 6, pp. 821–828, 2011.
- [85] L. Ali, I. Wajahat, N. A. Golilarz, F. Keshtkar, and S. A. C. Bukhari, "Lda-ga-svm: improved hepatocellular carcinoma prediction through dimensionality reduction and genetically optimized support vector machine," *Neural Computing and Applications*, vol. 33, no. 7, pp. 2783–2792, 2021.
- [86] S.-L. Zhu, J. Dong, C. Zhang, Y.-B. Huang, and W. Pan, "Application of machine learning in the diagnosis of gastric cancer based on noninvasive characteristics," *Plos one*, vol. 15, no. 12, p. e0244869, 2020.
- [87] M. A. Aslam, C. Xue, K. Wang, Y. Chen, A. Zhang, W. Cai, L. Ma, Y. Yang, X. Sun, M. Liu, *et al.*, "Svm based classification and prediction system for gastric cancer using dominant features of saliva," *Nano Biomed Eng*, vol. 12, no. 1, pp. 1–13, 2020.
- [88] Z. Song, S. Zou, W. Zhou, Y. Huang, L. Shao, J. Yuan, X. Gou, W. Jin, Z. Wang, X. Chen, *et al.*, "Clinically applicable histopathological diagnosis system for gastric cancer detection using deep learning," *Nature communications*, vol. 11, no. 1, pp. 1–9, 2020.
- [89] S. Umamaheswari, D. Sangeetha, C. Moulighanth, and E. Vignesh, "Kidnet: Kidney tumour diagnosis system design using deep convolutional neural network," in *Deep Learning Applications and Intelligent Decision Making in Engineering*, pp. 114–129, IGI Global, 2021.

- [90] S. Mascharak, B. J. Baird, and F. C. Holsinger, "Detecting oropharyngeal carcinoma using multispectral, narrow-band imaging and machine learning," *The Laryngoscope*, vol. 128, no. 11, pp. 2514–2520, 2018.
- [91] O. Kott, D. Linsley, A. Amin, A. Karagounis, C. Jeffers, D. Golijanin, T. Serre, and B. Gershman, "Development of a deep learning algorithm for the histopathologic diagnosis and gleason grading of prostate cancer biopsies: a pilot study," *European urology focus*, 2019.
- [92] K. Honda, Y. Hayashida, T. Umaki, T. Okusaka, T. Kosuge, S. Kikuchi, M. Endo, A. Tsuchida, T. Aoki, T. Itoi, *et al.*, "Possible detection of pancreatic cancer by plasma protein profiling," *Cancer research*, vol. 65, no. 22, pp. 10613–10622, 2005.
- [93] C. Baumgartner, C. Böhm, D. Baumgartner, G. Marini, K. Weinberger, B. Olgemöller, B. Liebl, and A. Roscher, "Supervised machine learning techniques for the classification of metabolic disorders in newborns," *Bioinformatics*, vol. 20, no. 17, pp. 2985–2996, 2004.
- [94] P. Danaee, R. Ghaeini, and D. A. Hendrix, "A deep learning approach for cancer detection and relevant gene identification," in *PACIFIC SYMPOSIUM ON BIOCOMPUTING 2017*, pp. 219–229, World Scientific, 2017.
- [95] M. M. Islam, M. R. Haque, H. Iqbal, M. M. Hasan, M. Hasan, and M. N. Kabir, "Breast cancer prediction: a comparative study using machine learning techniques," *SN Computer Science*, vol. 1, no. 5, pp. 1–14, 2020.
- [96] O. Klein, F. Kanter, H. Kulbe, P. Jank, C. Denkert, G. Nebrich, W. D. Schmitt, Z. Wu, C. A. Kunze, J. Sehouli, *et al.*, "Maldi-imaging for classification of epithelial ovarian cancer histotypes from a tissue microarray using machine learning methods," *PROTEOMICS—Clinical Applications*, vol. 13, no. 1, p. 1700181, 2019.
- [97] P. N. Yeganeh and M. T. Mostafavi, "Use of machine learning for diagnosis of cancer in ovarian tissues with a selected mrna panel," in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 2429–2434, 2018.
- [98] G. Wang, J. Y. C. Teoh, and K. S. Choi, "Diagnosis of prostate cancer in a chinese population by using machine learning methods," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1–4, 2018.
- [99] S. Liu, H. Zheng, Y. Feng, and W. Li, "Prostate cancer diagnosis using deep learning with 3d multi-parametric mri," in *Medical imaging 2017: computer-aided diagnosis*, vol. 10134, p. 1013428, International Society for Optics and Photonics, 2017.
- [100] N. Ito, H. Kawahira, H. Nakashima, M. Uesato, H. Miyauchi, and H. Matsubara, "Endoscopic diagnostic support system for ct1b colorectal cancer using deep learning," *Oncology*, vol. 96, no. 1, pp. 44–50, 2019.
- [101] Z. Zhou, D. Alvarez, C. Milla, and R. N. Zare, "Proof of concept for identifying cystic fibrosis from perspiration samples," *Proceedings of the National Academy of Sciences*, vol. 116, no. 49, pp. 24408–24412, 2019.
- [102] A. M. Abdeldaim, A. T. Sahlol, M. Elhoseny, and A. E. Hassanien, "Computer-aided acute lymphoblastic leukemia diagnosis system based on image analysis," in *Advances in Soft Computing and Machine Learning in Image Processing*, pp. 131–147, Springer, 2018.
- [103] L. Pan, G. Liu, F. Lin, S. Zhong, H. Xia, X. Sun, and H. Liang, "Machine learning applications for prediction of relapse in childhood acute lymphoblastic leukemia," *Scientific reports*, vol. 7, no. 1, pp. 1–9, 2017.
- [104] N. H. Alharbi, R. O. Bameer, S. S. Geddan, and H. M. Alharbi, "Recent advances and machine learning techniques on sickle cell disease," *Future Computing and Informatics Journal*, vol. 5, no. 1, p. 4, 2020.
- [105] S. Kolkur and D. Kalbande, "Survey of texture based feature extraction for skin disease detection," in *2016 International Conference on ICT in Business Industry & Government (ICTBIG)*, pp. 1–6, IEEE, 2016.
- [106] N. S. A. ALenezi, "A method of skin disease detection using image processing and machine learning," *Procedia Computer Science*, vol. 163, pp. 85–92, 2019.
- [107] S. Kuwayama, Y. Ayatsuka, D. Yanagisano, T. Uta, H. Usui, A. Kato, N. Takase, Y. Ogura, and T. Yasukawa, "Automated detection of macular diseases by optical coherence tomography and artificial intelligence machine learning of optical coherence tomography images," *Journal of ophthalmology*, vol. 2019, 2019.
- [108] Q. Zhao, K. Rosenbaum, R. Sze, D. Zand, M. Summar, and M. G. Linguraru, "Down syndrome detection from facial photographs using machine learning techniques," in *Medical Imaging 2013: Computer-Aided Diagnosis*, vol. 8670, p. 867003, International Society for Optics and Photonics, 2013.
- [109] A. Movaghar, M. Mailick, A. Sterling, J. Greenberg, and K. Saha, "Automated screening for fragile x premutation carriers based on linguistic and cognitive computational phenotypes," *Scientific reports*, vol. 7, no. 1, pp. 1–11, 2017.
- [110] F. Aszhari, Z. Rustam, F. Subroto, and A. Semendawai, "Classification of thalassemia data using random forest algorithm," in *Journal of Physics: Conference Series*, vol. 1490, p. 012050, IOP Publishing, 2020.
- [111] A. S. AlAgha, H. Faris, B. H. Hammo, and A.-Z. Ala'M, "Identifying β -thalassemia carriers using a data mining approach: The case of the gaza strip, palestine," *Artificial intelligence in medicine*, vol. 88,

- pp. 70–83, 2018.
- [112] H. Wen, Y. Liu, S. Wang, Z. Li, J. Zhang, Y. Peng, and H. He, "Multi-threshold white matter structural networks fusion for accurate diagnosis of tourette syndrome children," in *Medical Imaging 2017: Computer-Aided Diagnosis*, vol. 10134, p. 101341Q, International Society for Optics and Photonics, 2017.
 - [113] M. J. Morabito, M. Usta, X. Cheng, X. F. Zhang, A. Oztekin, and E. B. Webb, "Prediction of submonomer a2 domain dynamics of the von willebrand factor by machine learning algorithm and coarse-grained molecular dynamics simulation," *Scientific reports*, vol. 9, no. 1, pp. 1–11, 2019.
 - [114] G. Querin, M.-M. El Mendili, P. Bede, S. Delphine, T. Lenglet, V. Marchand-Pauvert, and P.-F. Pradat, "Multimodal spinal cord mri offers accurate diagnostic classification in als," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 89, no. 11, pp. 1220–1221, 2018.
 - [115] J. P. Felix, H. A. D. d. Nascimento, N. N. Guimarães, E. D. O. Pires, G. d. S. Vieira, and W. d. S. Alencar, "An effective and automatic method to aid the diagnosis of amyotrophic lateral sclerosis using one minute of gait signal," in *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 2745–2751, 2020.
 - [116] J. Venugopalan, L. Tong, H. R. Hassanzadeh, and M. D. Wang, "Multimodal deep learning models for early detection of alzheimer's disease stage," *Scientific reports*, vol. 11, no. 1, pp. 1–13, 2021.
 - [117] V. Saccà, A. Sarica, F. Novellino, S. Barone, T. Talarico, E. Filippelli, A. Granata, C. Chiriaco, R. B. Bossio, P. Valentino, *et al.*, "Evaluation of machine learning algorithms performance for the prediction of early multiple sclerosis from resting-state fmri connectivity data," *Brain imaging and behavior*, vol. 13, no. 4, pp. 1103–1114, 2019.
 - [118] J. Cai, F. Xing, A. Batra, F. Liu, G. A. Walter, K. Vandendorpe, and L. Yang, "Texture analysis for muscular dystrophy classification in mri with improved class activation mapping," *Pattern recognition*, vol. 86, pp. 368–375, 2019.
 - [119] K. Morita, A. Tashita, M. Nii, and S. Kobashi, "Computer-aided diagnosis system for rheumatoid arthritis using machine learning," in *2017 International Conference on Machine Learning and Cybernetics (ICMLC)*, vol. 2, pp. 357–360, 2017.
 - [120] J. Liu and N. Chen, "A 9 mrnas-based diagnostic signature for rheumatoid arthritis by integrating bioinformatic analysis and machine-learning," *Journal of Orthopaedic Surgery and Research*, vol. 16, no. 1, pp. 1–7, 2021.
 - [121] X. Huang, M. Khushi, M. Latt, C. Loy, and S. K. Poon, "Machine learning based method for huntington's disease gait pattern recognition," in *International Conference on Neural Information Processing*, pp. 607–614, Springer, 2019.
 - [122] J. Paula Felix, F. Henrique Teles Vieira, G. da Silva Vieira, R. Augusto Pereira Franco, R. Martins da Costa, and R. Lopes Salvini, "An automatic method for identifying huntington's disease using gait dynamics," in *2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI)*, pp. 1659–1663, 2019.
 - [123] K. D. Nguyen, L. A. Corben, P. N. Pathirana, M. K. Horne, M. B. Delatycki, and D. J. Szmulewicz, "Assessment of disease progression in friedreich ataxia using an instrumented self feeding activity," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 3827–3830, IEEE, 2020.
 - [124] Y. Moeuf, O. Phalla, M. Gouysse, N. Pasi, M. Tchitchinadze, O. Milleron, F. Arnoult, S. Schmitt, G. Jondeau, and C. Bouleti, "Dural ectasia re-definition in marfan syndrome," *Archives of Cardiovascular Diseases Supplements*, vol. 11, no. 1, p. 57, 2019.
 - [125] Y. Zhu, Z.-g. Liu, L. Yang, K. Wang, M.-H. Wang, X.-y. Yang, X. Wu, X. Tian, R.-G. Zhang, J.-c. Ma, *et al.*, "Development and validation of a deep learning model for preoperative screening of myasthenia gravis in patients with thymoma based on ct images," 2020.
 - [126] G. Pahuja and T. Nagabhushan, "A comparative study of existing machine learning approaches for parkinson's disease detection," *IETE Journal of Research*, vol. 67, no. 1, pp. 4–14, 2021.
 - [127] S. Khan, F. Azam, M. W. Anwar, Y. Rasheed, M. Saleem, and N. Ejaz, "A novel data mining approach for detection of polio disease using spatio-temporal analysis," in *Proceedings of the 2020 6th International Conference on Computing and Artificial Intelligence*, pp. 223–227, 2020.
 - [128] S. R. Kerketta and D. Ghosh, "Detection of onset and progression of osteoporosis using machine learning," *Machine Learning for Healthcare Applications*, p. 137, 2021.
 - [129] E. Vaghefi, S. Hill, H. M. Kersten, and D. Squirrell, "Multimodal retinal image analysis via deep learning for the diagnosis of intermediate dry age-related macular degeneration: a feasibility study," *Journal of ophthalmology*, vol. 2020, 2020.
 - [130] S. Solis-Reyes, M. Avino, A. Poon, and L. Kari, "An open-source k-mer based machine learning tool for fast and accurate subtyping of hiv-1 genomes," *PloS one*, vol. 13, no. 11, p. e0206409, 2018.
 - [131] H. R. Elder, S. Gruber, S. J. Willis, N. Cocoros, M. Callahan, E. W. Flagg, M. Klompas, and K. K. Hsu, "Can machine learning help identify patients at risk for recurrent sexually transmitted infections?," *Sexually Transmitted Diseases*, vol. 48, no. 1, pp. 56–62, 2021.

- [132] S. Klein, A. Quaas, J. Quantius, H. Löser, J. Meinel, M. Peifer, S. Wagner, S. Gattenlöhner, C. Wittekindt, M. von Knebel Doeberitz, *et al.*, "Deep learning predicts hpv association in oropharyngeal squamous cell carcinomas and identifies patients with a favorable prognosis using regular h&e stains," *Clinical Cancer Research*, vol. 27, no. 4, pp. 1131–1138, 2021.
- [133] M. Hussain, S. Zhu, P. Yang, Y. An, Z. Li, I. Ali, B. Liu, H. Shen, and N. He, "Rapid detection system for hepatitis b surface antigen (hbsag) based on immunomagnetic separation, multi-angle dynamic light scattering and support vector machine," *IEEE Access*, vol. 8, pp. 107373–107386, 2020.
- [134] H.-Y. Wang, C.-C. Hung, C.-H. Chen, T.-Y. Lee, K.-Y. Huang, H.-C. Ning, N.-C. Lai, M.-H. Tsai, L.-C. Lu, Y.-J. Tseng, *et al.*, "Increase trichomonas vaginalis detection based on urine routine analysis through a machine learning approach," *Scientific reports*, vol. 9, no. 1, pp. 1–10, 2019.
- [135] W. Rahane, H. Dalvi, Y. Magar, A. Kalane, and S. Jondhale, "Lung cancer detection using image processing and machine learning healthcare," in *2018 International Conference on Current Trends towards Converging Technologies (ICCTCT)*, pp. 1–5, 2018.
- [136] S. Makaju, P. Prasad, A. Alsadoon, A. Singh, and A. Elchouemi, "Lung cancer detection using ct scan images," *Procedia Computer Science*, vol. 125, pp. 107–114, 2018.
- [137] I. M. Nasser and S. S. Abu-Naser, "Lung cancer detection using artificial neural network," *International Journal of Engineering and Information Systems (IJEAIS)*, vol. 3, no. 3, pp. 17–23, 2019.
- [138] M. Schwyzer, D. A. Ferraro, U. J. Muehlemaier, A. Curioni-Fontecedro, M. W. Huellner, G. K. Von Schulthess, P. A. Kaufmann, I. A. Burger, and M. Messerli, "Automated detection of lung cancer at ultralow dose pet/ct by deep neural networks—initial results," *Lung Cancer*, vol. 126, pp. 170–173, 2018.
- [139] W. Sun, B. Zheng, and W. Qian, "Computer aided lung cancer diagnosis with deep learning algorithms," in *Medical imaging 2016: computer-aided diagnosis*, vol. 9785, p. 97850Z, International Society for Optics and Photonics, 2016.
- [140] K. Jindal, N. Baliyan, and P. S. Rana, "Obesity prediction using ensemble machine learning approaches," in *Recent Findings in Intelligent Computing Techniques*, pp. 355–362, Springer, 2018.
- [141] M. R. Lee, V. Sankar, A. Hammer, W. G. Kennedy, J. J. Barb, P. G. McQueen, and L. Leggio, "Using machine learning to classify individuals with alcohol use disorder based on treatment seeking status," *EClinicalMedicine*, vol. 12, pp. 70–78, 2019.
- [142] S. Paul, S. K. Jandhyala, and T. Basu, "Early detection of signs of anorexia and depression over social media using effective machine learning frameworks," in *CLEF (Working notes)*, 2018.
- [143] D. Raab, H. Baumgartl, and R. Buettner, "Machine learning based diagnosis of binge eating disorder using eeg recordings," in *PACIS*, p. 97, 2020.
- [144] S. of Nuclear Medicine and M. Imaging, "Types of brain disorders," 2020.
- [145] E. Healthcare, "Tests and diagnostics," 2019.
- [146] M. N. Today, "Everything you need to know about heart disease," 2020.
- [147] M. Clinic, "Heart disease- diagnosis and treatment," 2021.
- [148] M. Plus, "Lung disease," 2021.
- [149] A. L. Association, "Lung procedures, tests, & treatments," 2021.
- [150] S. Johnson and C. Stephens, "Kidney health and kidney disease basics," 2018.
- [151] M. Clinic, "Chronic kidney disease," 2021.
- [152] K. Brind'Amour, "All about common skin disorders," 2016.
- [153] M. Plus, "Skin conditions," 2020.
- [154] M. U. of South Carolina, "Diseases of the stomach & duodenum," 2020.
- [155] J. H. Medicine, "Diagnosis and screening for gastric conditions," 2021.
- [156] S. W. . S. Sampson, "Autoimmune diseases: Types, symptoms, causes, and more," 2019.
- [157] J. H. Medicine, "What are common symptoms of autoimmune disease?," 2021.
- [158] M. Plus, "Mental disorders," 2019.
- [159] M. Clinic, "Mental illness," 2021.
- [160] S. B. Kotsiantis, D. Kanellopoulos, and P. E. Pintelas, "Data preprocessing for supervised learning," *International Journal of Computer Science*, vol. 1, no. 2, pp. 111–117, 2006.
- [161] S. Dreiseitl and L. Ohno-Machado, "Logistic regression and artificial neural network classification models: a methodology review," *Journal of biomedical informatics*, vol. 35, no. 5-6, pp. 352–359, 2002.
- [162] A. Singh, N. Thakur, and A. Sharma, "A review of supervised machine learning algorithms," in *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, pp. 1310–1315, 2016.
- [163] C. Zhang, J. Zhao, F. Li, H. Jia, and J. Tian, "Support vector machine algorithm for real-time detection of vf signals," *Procedia Environmental Sciences*, vol. 8, pp. 602–608, 2011.
- [164] M. Awad and R. Khanna, "Support vector regression," in *Efficient learning machines*, pp. 67–80, Springer, 2015.
- [165] "Quick guide to evaluation metrics for supervised and unsupervised machine learning," 2020.
- [166] L. R. Borges, "Analysis of the wisconsin breast cancer dataset and machine learning for breast cancer detection," *Group*, vol. 1, no. 369, pp. 15–19, 1989.

- [167] J. Forsström, P. Nuutila, and K. Irjala, "Using the id3 algorithm to find discrepant diagnoses from laboratory databases of thyroid patients," *Medical Decision Making*, vol. 11, no. 3, pp. 171–175, 1991.
- [168] W. N. Street, "Cancer diagnosis and prognosis via linear-programming-based machine learning," tech. rep., 1994.
- [169] W. R. Shankle, S. Mani, M. J. Pazzani, and P. Smyth, "Detecting very early stages of dementia from normal aging with machine learning methods," in *Conference on artificial intelligence in medicine in Europe*, pp. 71–85, Springer, 1997.
- [170] M. Kukar, I. Kononenko, et al., "Cost-sensitive learning with neural networks.," in *ECAI*, vol. 15, pp. 88–94, Citeseer, 1998.
- [171] J. Sepulveda-Sanchis, G. Camps-Valls, E. Soria-Olivas, S. Salcedo-Sanz, C. Bousoño-Calzon, G. Sanz-Romero, and J. Marrugat de la Iglesia, "Support vector machines and genetic algorithms for detecting unstable angina," in *Computers in Cardiology*, pp. 413–416, 2002.
- [172] M. Clinic, "Ecg," 2021.
- [173] Y. Liu, "Active learning with support vector machine applied to gene expression data for cancer classification," *Journal of chemical information and computer sciences*, vol. 44, no. 6, pp. 1936–1941, 2004.
- [174] M. Crow, N. Lim, S. Ballouz, P. Pavlidis, and J. Gillis, "Predictability of human differential gene expression," *Proceedings of the National Academy of Sciences*, vol. 116, no. 13, pp. 6491–6500, 2019.
- [175] H. Selvaraj, S. T. Selvi, D. Selvathi, and L. Gewali, "Brain mri slices classification using least squares support vector machine," *International Journal of Intelligent Computing in Medical Sciences & Image Processing*, vol. 1, no. 1, pp. 21–33, 2007.
- [176] M. Clinic, "Mri," 2021.
- [177] W. Yu, T. Liu, R. Valdez, M. Gwinn, and M. J. Khoury, "Application of support vector machine modeling for prediction of common diseases: the case of diabetes and pre-diabetes," *BMC medical informatics and decision making*, vol. 10, no. 1, pp. 1–7, 2010.
- [178] T. V. Sriram, M. V. Rao, G. S. Narayana, D. Kaladhar, and T. P. R. Vital, "Intelligent parkinson disease prediction using machine learning algorithms," *Int. J. Eng. Innov. Technol*, vol. 3, pp. 212–215, 2013.
- [179] N. Designer, "Telemonitoring of parkinson's disease progression by non-invasive speech tests," 2020.
- [180] N. M. Tahir and H. H. Manap, "Parkinson disease gait classification based on machine learning approach," *Journal of Applied Sciences(Faisalabad)*, vol. 12, no. 2, pp. 180–185, 2012.
- [181] SIMPLIFASTER, "Hawkin dynamics dual force plates," 2021.
- [182] S. Potluri, S. Ravuri, C. Diedrich, and L. Schega, "Deep learning based gait abnormality detection using wearable sensor system," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 3613–3619, IEEE, 2019.
- [183] A. Al-Zebari and A. Sengur, "Performance comparison of machine learning techniques on diabetes disease detection," in *2019 1st International Informatics and Software Engineering Conference (UBMYK)*, pp. 1–4, IEEE, 2019.
- [184] A. Imran, I. Posokhova, H. N. Qureshi, U. Masood, S. Riaz, K. Ali, C. N. John, and M. Nabeel, "Ai4covid-19: Ai enabled preliminary diagnosis for covid-19 from cough samples via an app," *arXiv preprint arXiv:2004.01275*, 2020.
- [185] R. Das, I. Turkoglu, and A. Sengur, "Diagnosis of valvular heart disease through neural networks ensembles," *Computer methods and programs in biomedicine*, vol. 93, no. 2, pp. 185–191, 2009.
- [186] Centers for Disease Control and Prevention, "Nhanes questionnaires, datasets, and related documentation."
- [187] "Framingham heart study," 2020.
- [188] S. Namuduri, B. N. Narayanan, V. S. P. Davuluru, L. Burton, and S. Bhansali, "Deep learning methods for sensor based predictive maintenance and future perspectives for electrochemical sensors," *Journal of The Electrochemical Society*, vol. 167, no. 3, p. 037552, 2020.
- [189] CDC, "Obesity and overweight," 2021.
- [190] M. Clinic, "Obesity," 2021.
- [191] CDC, "What are the risk factors for lung cancer," 2020.
- [192] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. E. Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," *Physical Review E*, vol. 64, no. 6, p. 061907, 2001.
- [193] A. Swetha, "Eeg dataset of epileptic seizure patients," 2021.
- [194] N. Stevenson, K. Tapani, L. Lauronen, and S. Vanhatalo, "A dataset of neonatal eeg recordings with seizure annotations," *Scientific data*, vol. 6, no. 1, pp. 1–8, 2019.
- [195] "Chronic kidney disease dataset." https://archive.ics.uci.edu/ml/datasets/chronic_kidney_disease. 2021 – 09 – 30.
- [196] S. K. Singh, A. Malik, A. Firoz, and V. Jha, "Cdkd: a clinical database of kidney diseases," *BMC nephrology*, vol. 13, no. 1, pp. 1–10, 2012.
- [197] F. Thabtah, "Autism spectrum disorder screening: machine learning adaptation and dsm-5 fulfillment," in *Proceedings of the 1st International Conference on Medical and health Informatics 2017*, pp. 1–6, 2017.
- [198] E. Billing, T. Belpaeme, H. Cai, H.-L. Cao, A. Ciocan,

- C. Costescu, D. David, R. Homewood, D. Hernandez Garcia, P. Gómez Esteban, *et al.*, "The dream dataset: Supporting a data-driven study of autism spectrum disorder and robot enhanced therapy," *PloS one*, vol. 15, no. 8, p. e0236939, 2020.
- [199] F. Qureshi, J. Adams, D. Coleman, D. Quig, and J. Hahn, "Urinary essential elements of young children with autism spectrum disorder and their mothers," *Research in autism spectrum disorders*, vol. 72, p. 101518, 2020.
- [200] "Autism brain imaging data exchange." https://fcon_1000.projects.nitrc.org/indi/abide/. Accessed: 2021-10-05.
- [201] E. Garcia-Ceja, M. Riegler, P. Jakobsen, J. Tørresen, T. Nordgreen, K. J. Oedegaard, and O. B. Fasmer, "Depresjon: a motor activity database of depression episodes in unipolar and bipolar patients," in *Proceedings of the 9th ACM multimedia systems conference*, pp. 472–477, 2018.
- [202] "Coswara-data." <https://github.com/iiscleap/Coswara-Data>. Accessed: 2021-10-31.
- [203] "Ovation COVID-19 diagnostic test results (US)." <https://aws.amazon.com/marketplace/pp/prodview-uqtrgyqowwoheoffers>. Accessed: 2021-10-31.
- [204] "Extensive COVID-19 x-ray and CT chest images dataset." <https://data.mendeley.com/datasets/8h65ywd2jr/3>. Accessed: 2021-10-31.
- [205] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, *et al.*, "The multimodal brain tumor image segmentation benchmark (brats)," *IEEE transactions on medical imaging*, vol. 34, no. 10, pp. 1993–2024, 2014.
- [206] S. Bakas, M. Reyes, A. Jakab, S. Bauer, M. Rempfler, A. Crimi, R. T. Shinohara, C. Berger, S. M. Ha, M. Rozycki, *et al.*, "Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge," *arXiv preprint arXiv:1811.02629*, 2018.
- [207] F. Prior, K. Smith, A. Sharma, J. Kirby, L. Tarbox, K. Clark, W. Bennett, T. Nolan, and J. Freymann, "The public cancer radiology imaging collections of the cancer imaging archive," *Scientific data*, vol. 4, no. 1, pp. 1–7, 2017.
- [208] K. Schmainda and M. Prah, "Data from brain-tumor-progression," *The Cancer Imaging Archive*, 2018.
- [209] H. Ali, M. Yasmin, M. Sharif, and M. H. Rehmani, "Computer assisted gastric abnormalities detection using hybrid texture descriptors for chromoendoscopy images," *Computer methods and programs in biomedicine*, vol. 157, pp. 39–47, 2018.
- [210] Y. Sakai, S. Takemoto, K. Hori, M. Nishimura, H. Ikematsu, T. Yano, and H. Yokota, "Automatic detection of early gastric cancer in endoscopic images using a transferring convolutional neural network," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 4138–4141, IEEE, 2018.
- [211] H. J. Yoon, S. Kim, J.-H. Kim, J.-S. Keum, S.-I. Oh, J. Jo, J. Chun, Y. H. Youn, H. Park, I. G. Kwon, *et al.*, "A lesion-based convolutional neural network improves endoscopic detection and depth prediction of early gastric cancer," *Journal of clinical medicine*, vol. 8, no. 9, p. 1310, 2019.
- [212] "Heart disease data set." <https://archive.ics.uci.edu/ml/datasets/heart+disease>. Accessed: 2021-09-30.
- [213] T. Ahmad, A. Munir, S. H. Bhatti, M. Aftab, and M. A. Raza, "Survival analysis of heart failure patients: A case study," *PloS one*, vol. 12, no. 7, p. e0181001, 2017.
- [214] "Chronic kidney disease dataset." Heart Attack Risk Prediction Using Machine Learning. Accessed: 2021-09-30.
- [215] "Heart statlog, howpublished = <https://datahub.io/machine-learning/heart-statlog#readme>, note = Accessed: 2021-09-30."
- [216] "Uci diabetes dataset." <https://archive.ics.uci.edu/ml/datasets/diabetes>. Accessed: 2021-09-30.
- [217] "Pima indians diabetes database." <https://www.kaggle.com/uciml/pima-indians-diabetes-database>. Accessed: 2021-09-30.
- [218] "Alzheimer's disease neuroimaging initiative." <http://adni.loni.usc.edu/>. Accessed: 2021-10-11.
- [219] "Alzheimer's disease and healthy aging data portal." <https://chronicdata.cdc.gov/Healthy-Aging/Alzheimer-s-Disease-and-Healthy-Aging-Data/hfr9-rurv>. Accessed: 2021-10-11.
- [220] L. K. Leong and A. A. Abdullah, "Prediction of alzheimer's disease (ad) using machine learning techniques with boruta algorithm as feature selection method," in *Journal of Physics: Conference Series*, vol. 1372, p. 012065, IOP Publishing, 2019.
- [221] J. H. Park, H. E. Cho, J. H. Kim, M. M. Wall, Y. Stern, H. Lim, S. Yoo, H. S. Kim, and J. Cha, "Machine learning prediction of incidence of alzheimer's disease using large-scale administrative health data," *NPJ digital medicine*, vol. 3, no. 1, pp. 1–7, 2020.
- [222] T. Rögnvaldsson, L. You, and D. Garwicz, "State of the art prediction of hiv-1 protease cleavage sites," *Bioinformatics*, vol. 31, no. 8, pp. 1204–1210, 2015.
- [223] E. Orel, R. Esra, J. Estill, S. Marchand-Maillet, A. Merzouki, and O. Keiser, "Machine learning to identify socio-behavioural predictors of hiv positivity in east and southern africa," *medRxiv*, 2020.
- [224] K. Fernandes, J. S. Cardoso, and J. Fernandes, "Transfer learning with partial observability applied to cervical cancer screening," in *Iberian conference on*

- pattern recognition and image analysis*, pp. 243–250, Springer, 2017.
- [225] S. A. Thamrin, D. S. Arsyad, H. Kuswanto, A. Lawi, and S. Nasir, “Predicting obesity in adults using machine learning techniques: An analysis of indonesian basic health research 2018,” *Frontiers in Nutrition*, vol. 8, p. 252, 2021.
- [226] F. M. Palechor and A. de la Hoz Manotas, “Dataset for estimation of obesity levels based on eating habits and physical condition in individuals from colombia, peru and mexico,” *Data in brief*, vol. 25, p. 104344, 2019.
- [227] E. De-La-Hoz-Correa, F. Mendoza Palechor, A. De-La-Hoz-Manotas, R. Morales Ortega, and A. B. Sánchez Hernández, “Obesity level estimation software based on decision trees,” 2019.
- [228] “A large-scale ct and pet/ct dataset for lung cancer diagnosis (lung-pet-ct-dx).” <https://tinyurl.com/kn4s73zp>. Accessed: 2021-10-11.
- [229] “Nelson digital data for lung cancer.” <https://www.umcg.nl/EN/corporate/Departments/nelson/dataset/Pages/default.aspx>. Accessed: 2021-10-11.
- [230] A. F. Markus, J. A. Kors, and P. R. Rijnbeek, “The role of explainability in creating trustworthy artificial intelligence for health care: a comprehensive survey of the terminology, design choices, and evaluation strategies,” *Journal of Biomedical Informatics*, p. 103655, 2020.
- [231] K. Ahmad, M. Maabreh, M. Ghaly, K. Khan, J. Qadir, and A. Al-Fuqaha, “Developing future human-centered smart cities: Critical analysis of smart city security, interpretability, and ethical challenges,” *arXiv preprint arXiv:2012.09110*, 2020.
- [232] A. Qayyum, A. Ijaz, M. Usama, W. Iqbal, J. Qadir, Y. Elkhatib, and A. Al-Fuqaha, “Securing machine learning in the cloud: A systematic review of cloud machine learning security,” *Frontiers in big Data*, vol. 3, 2020.
- [233] H. Kasyap and S. Tripathy, “Privacy-preserving decentralized learning framework for healthcare system,” *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 17, no. 2s, pp. 1–24, 2021.
- [234] K. H. Nam, D. H. Kim, B. K. Choi, and I. H. Han, “Internet of things, digital biomarker, and artificial intelligence in spine: Current and future perspectives,” *Neurospine*, vol. 16, no. 4, p. 705, 2019.
- [235] M. Poduval, A. Ghose, S. Manchanda, V. Bagaria, and A. Sinha, “Artificial intelligence and machine learning: A new disruptive force in orthopaedics,” *Indian Journal of Orthopaedics*, pp. 1–14, 2020.
- [236] “Meeting the challenges of ai in health care.” <https://bmcmmedicine.biomedcentral.com/articles/10.1186/s12916-019-1426-2>. Accessed: 2021-11-01.
- [237] C. J. L. Murray and et al., “Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the global burden of disease study 2019,” vol. 396, no. 10258, pp. 1223–1249. Publisher: Elsevier.
- [238] R. Guo, L. Cheng, J. Li, P. R. Hahn, and H. Liu, “A survey of learning causality with data: Problems and methods,” vol. 53, no. 4, pp. 75:1–75:37.
- [239] P. Lecca, “Machine learning for causal inference in biological networks: Perspectives of this challenge,” vol. 1, p. 45.
- [240] J. G. Richens, C. M. Lee, and S. Johri, “Improving the accuracy of medical diagnosis with causal machine learning,” vol. 11, no. 1, p. 3923. Bandiera_abtest: a Cc_license_type: cc_by Cg_type: Nature Research Journals Number: 1 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Diagnosis;Predictive medicine;Statistics Subject_term_id: diagnosis;predictive-medicine;statistics.