

Machine Learning Approaches for Precision Medicine: A Review

Angga Aditya Permana¹, Maureen Zerlina Oktaviani¹, Analekta Tiara Perdana²

¹Department of Informatics, Faculty of Engineering and Informatics, Universitas Multimedia Nusantara, Banten, Indonesia

²Department of Biology, Faculty of Science, UIN Sultan Maulana Hasanuddin, Banten, Indonesia

angga.permana@umn.ac.id (Corresponding Author), maureen.zerlina @student.umn.ac.id, analekta.tiara@uinbanten.ac.id

Abstract. Despite the development of numerous disease treatments, the need for individualized patient-specific treatments remains. Precision medicine aims to identify the most suitable treatment for patients by tailoring it to each patient's unique characteristics. Machine learning (ML) approaches are increasingly being utilized to analyze large datasets and identify patterns that can help predict disease risk, diagnosis, and treatment outcomes. In this systematic literature review, the authors collected and analyzed 37 articles that discuss the application of ML approaches for precision medicine. The review reveals that Random Forest (RF) and Support Vector Machine (SVM) are commonly utilized ML algorithms for precision medicine, and classification is the most frequently used type of ML between 2013 and 2022. The authors hope that the findings of this systematic literature review will aid researchers in developing ML-based precision medicine studies.

Keywords: Machine learning, precision medicine, treatment, systematic literature review

1. Introduction

Health is one of the most valuable human assets, and a serious disease is one of life's most troublesome trials. Even if there is a hope for a full recovery, it is very difficult to be patient, there must be an urge to worry about doing something to try to get well (Gard et al., 2013). According to the World Health Organization (WHO), the leading causes of death worldwide are ischemic heart disease and stroke. Other serious disease that can cause death are chronic obstructive pulmonary disease (COPD), lower respiratory tract infections, Alzheimer's disease, and diabetes. Meanwhile according to CDC, Indonesian people face cardiovascular disease, neoplasms, diabetes & kidney diseases, digestive diseases, and respiratory infections as other serious illnesses that lead to death. To minimize this problem, a more up-to-date treatment technique is needed compared to what currently exists (Masic et al., 2008).

Precision medicine is a method of modern medicine that aims to develop prevention and treatment strategies based on a combination of environment, lifestyle, genetics, and individual phenotypic diversity (Clish, 2015). Another goal of precision medicine is to provide tailored medical treatment for patient according to their genetic characteristics (Ho et al., 2019). This approach is much more effective against the disease than medicine which treats everyone in the same way (Akben, 2018; Khurana Hershey et al., 2022).

Machine learning (ML) is designed as a unique transformative method to discover hidden biological interactions to better predict and diagnose complex disease (Gaudillo et al., 2019). A subfield of data science known as "machine learning" teaches computers to perform tasks by analyzing patterns in large datasets and developing rules or algorithms that maximize task performance. (Glazyrin et al., 2020). Machine learning algorithms are commonly used in many fields such as big data, internet of things, social networking, cybersecurity, computer-aided diagnosis of diseases, analysis of biomedical data, and intrusion detection (Dritsas & Trigka, 2022; Nasarian et al., 2020). In biological field in particular, ML assist for example by coding region recognition, prediction of signal peptides, identification of biomarkers, recognition of disease genes, detection of metabolic network, and protein-protein interactions (Korani et al., 2019). ML holds an important role for disease diagnosis by simply analyzing existing patient records and training model to predict new patient behaviors (Ilyas et al., 2021).

ML as an approaches of precision medicine will be a perfect combination and very potential to improve patient healthcare. Advanced computation that will be provided by ML enabling diagnostic system that will enhance decision making of the physician. This review article is concentrated on the contribution of ML as an approaches of precision medicine. The learning ability of ML can be utilized to optimize measurement performance of biomedical data analysis, especially determination and anticipation of patient disease risk (Quazi, 2022).

There have been many related studies discussing the topic of developing precision medicine for several diseases using a machine learning approach. For examples (Ma et al., 2020), (da Silveira et al., 2022; Hayes et al., 2021; N et al., 2019) are predicting and diagnosing CKD by using the developed algorithms which are HMANN, elastic net regression, ABC4.5, DT, and RF. Then (Ed-daoudy & Maalmi, 2020) classified breast cancer by using association rules and SVM. (López et al., 2018) succeeded in researching the RF technique in dealing with the complexity of SNP datasets used to predict T2D disease. Research (B. Yu et al., 2021) aims to predict protein-protein interactions using elastic net and deep forest algorithms. Research (Blair et al., 2014; Burnard et al., 2022; Hess et al., 2017; Liu & Xuan, 2015; F. Yu et al., 2014) discuss about the well-known biomarker, SNP, with machine learning approaches. Artificial Neural Network (ANN) algorithm is applied on (Ghafouri-Fard et al., 2019, 2020) researches for diagnosing autism spectrum disorder (ASD) and predicting multiple sclerosis disease. Not only used for the benefit of disease in humans, it is also used for the benefit of other living things, namely animals and plants such as research (Bradley et al., 2019; Korani et al., 2019; Yao et al., 2013). Several research references regarding the comparison of the application of machine learning in terms of precision medicine, such as in studies (Almansour et al., 2019; Antony et al., 2021; Kao et al., 2022; Kucukkal et al., 2014; Lee et al., 2022; Xiao et al., 2019; Zubair Hasan & Zahid Hasan, 2019) were also selected for review.

This systematic literature review aims to identify and analyze several literatures published between 2013-2022 to determine what can be concluded about machine learning approaches for precision medicine, such

as search keywords, and the algorithms and types of machine learning are used.

2. Literature Review

2.1. Systematic literature review

A systematic literature review is an action of recognizing, evaluating, and interpreting the distribution exploration articles that are relevant to the research question being carried out. Systematic literature review can also be conducted to examine the extent to which empirical evidence supports or contradicts theoretical hypothesis, or even to assist with creating new theory (Carvalho et al., 2019).

2.2. Precision medicine

Precision medicine is a modern concept that has been used to describe accurate medical treatment tailored to the individual characteristics of each patient, in which the patient's disease is analyzed according to the molecular data, genomics, and systems biology models to define the patient's disease at the molecular level and choose the right treatment. Precision medicine promises to combine individual data about genetic susceptibility to disease, biomarker data on disease risk and response, also physiological and behavioral data from sensors and new databases to create a rich model that can predicts risk and response with high accuracy. Most scientists are increasingly using the term 'precision medicine' in light of the utilization of genomics to characterizing sicknesses for individualized therapy. Using genetic information and other databases from a big data perspective, diabetes and other disease are classified into disease subgroups, each with its own best treatment (Klonoff, 2015).

2.3. Machine learning

Machine Learning (ML) is a branch of Artificial Intelligence (AI) with a focus on computers that can make decisions and learn automatically [12]. In ML, algorithm uses the ability to distinguish features to classify samples. The performance of a ML model depends not only on the particular training algorithms, but also depends on the properties of the input data, such as the features number and correlation between the features (Ebiaredoh-Mienye et al., 2022). Furthermore, machine learning algorithms are divided into supervised learning, unsupervised learning, and reinforcement learning. Supervised learning is well known for its classification and regression methods, where sample datasets fully labeled to classify the unknown classes [39] (Senan et al., 2021). In contrast to that, unsupervised learning defines a function by describing the hidden structure of data that has no labels. Unsupervised learning can be categorized into clustering and association. Lastly, reinforcement learning has a fundamental to learn the experience or data with trial and error. Algorithms on reinforcement learning related to dynamic programming algorithm (Korani et al., 2019).

3. Research Method

3.1. Research question

1. RQ-1: Which journals publish articles discussing machine learning approaches to precision medicine?
2. RQ-2: What are the keywords in the machine learning approach research for precision medicine?
3. RQ-3: What algorithms are applied in the machine learning approach research for precision medicine?
4. RQ-4: What types of machine learning are used in machine learning approach research for precision medicine?

3.2. Record identification and screening

3.2.1 Keyword decision

The use of the right keywords is very necessary to collect articles that are in accordance with the discussed research topics. In this research, keywords are collected by identifying the existing terms on research questions, the use of root words and affixes according to the research theme, the use of 'AND' and 'OR' operators according to needs.

This research reviews articles that indexed by Scopus, so the keywords used are in English. The search

strings are obtained as follows:

- “Chronic Kidney Disease” AND “machine learning”
- (“Single Nucleotide Polymorphism” OR “SNP”) AND “Machine Learning”
- (“Single Nucleotide Polymorphism” OR “SNP”) AND “Elastic Net”
- (“Kidney” OR “Renal”) AND “Elastic Net”

The search strings above are used as search keywords for articles that used in this study. Articles search were carried out using the Harzing’s Publish or Perish (PoP) applications and filtered only articles indexed by Scopus. For each keyword, obtained 200 articles that related to the research topic, so it is equal to 800 articles but only 37 articles are eligible for this research. The graph of the distribution of the number of articles based on the Scopus Quartile Rating is shown in the following figure.

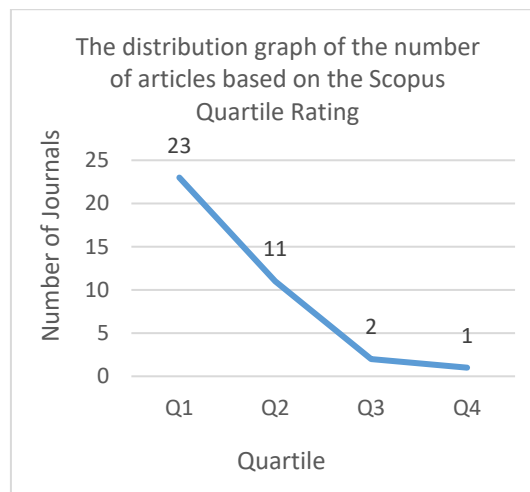


Fig. 1: The distribution graph of the number of published articles based on the Scopus quartile rating

3.2.2 Quality evaluation

1. Inclusion exclusion

To get articles that are most relevant to the research topic, it is necessary to select criteria consisting of inclusion criteria and exclusion criteria. Inclusion criteria are the criteria used as criteria in selecting literature, while exclusion criteria are not used in selecting literature.

Table 1: Inclusion Criteria and Exclusion Criteria

No	Inclusion Criteria	Exclusion Criteria
1	Research journals published between 2013-2022	Research journals published before 2013
2	Research journals are in English	Research journals are not in English
3	Research journals are indexed by Scopus	Research journals are not indexed by Scopus
4	Research journals discussing about machine learning for medicine precision	Research journals discussing about medicine precision not including machine learning approaches

2. PRISMA Flowchart

This research adapts recommendation outline in Preferred Reporting Items for Systematic Reviews and Meta-Analyses or known as PRISMA statement. Information sources are obtained from Scopus by

searching using the keywords in section “Keywords decision”. The collected articles are articles published in 2013 – 2022. Then, read and filter the articles through abstract, determine which articles are worthy of further review. In fact, there are only 39 articles were selected due to its eligibility and suitability, and finally there are 37 articles were used in this review research.

This research work gathered 800 articles about machine learning approach for precision medicine. Upon further research reveals there are only 46 related articles on this topic. Furthermore, 37 articles were selected to be discussed and presented in this review article. As such, it can be discovered which journals have published the most articles on machine learning approaches for precision medicine, which keywords the authors used most frequently when researching this topic, and what algorithms and types of machine learning approaches are most widely used in research related to this precision medicine.

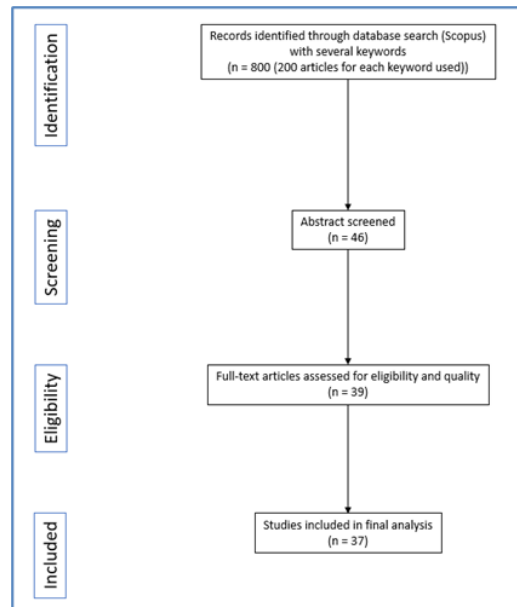


Fig. 2: PRISMA flow diagram for the systematic review of articles included in this review

4. Result and Discussion

4.1. Research article based on journal publisher

This research reviewed journal that been published by several publishers. Totally there are 42 different journals that publish articles about machine learning approaches for precision medicine and used for this research. Elsevier and MDPI with different research area contribute by giving 6 related research articles.

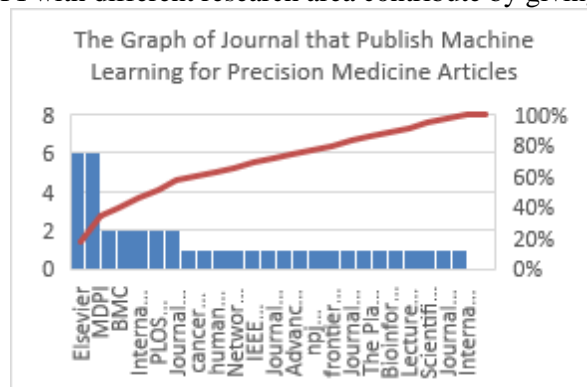


Fig. 3: The Journal Graph that Publish Articles containing Machine Learning Approaches for Precision Medicine

4.2. Keywords

Keywords are necessary to facilitate the research. Based on the literature that used in this research, the most used keywords on machine learning for precision medicine are *machine learning*, *chronic kidney disease*, and *SNP*.

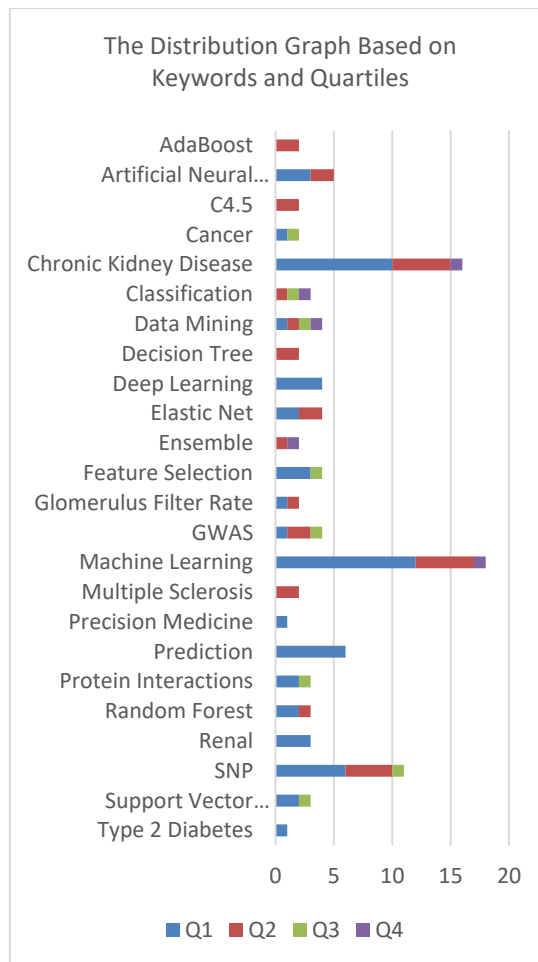


Fig. 4: The Distribution Graph Based on Keywords and Quartiles

4.3. Machine learning algorithms

Finding the correct machine learning algorithm is necessary for this research to be used as material for comparison which algorithm can provide the highest accuracy. According to the distribution graph below, shown that random forest is the most preferred algorithm, followed by support vector machine algorithm.

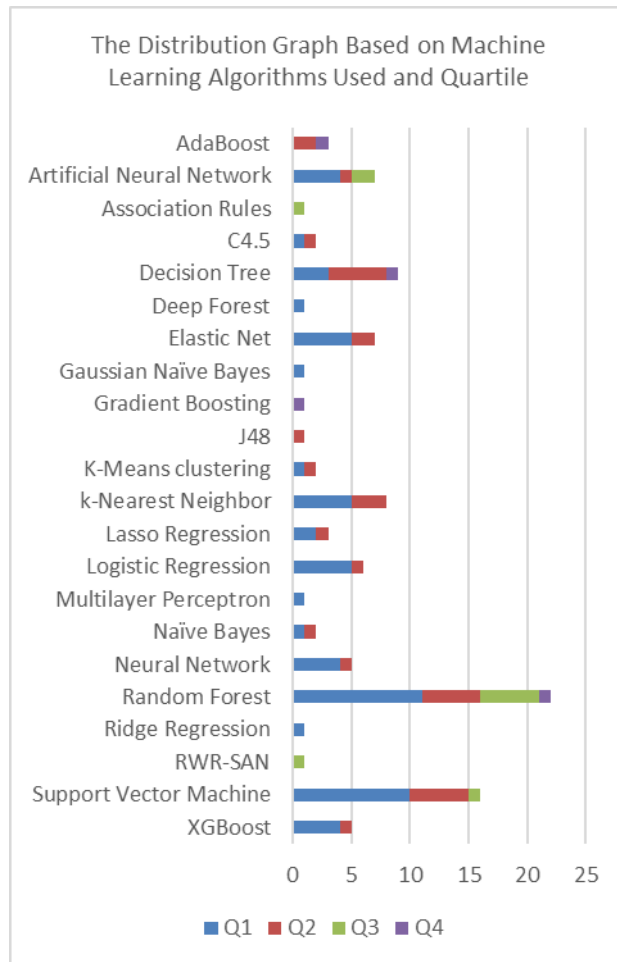


Fig. 5: The Distribution Graph of Machine Learning Algorithms

4.4. Machine learning types

Classification is the most used type of machine learning for articles related to machine learning approaches for precision medicine, followed by regression and ensemble types of machine learning.

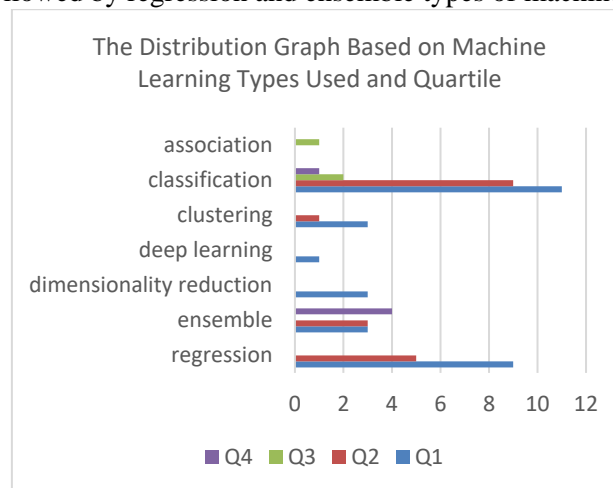


Fig. 6: The Distribution Graph of Machine Learning Types

Based on the 37 research articles that have been selected and reviewed using the PRISMA statement, the following table is the conclusion of these articles.

Table 2 Conclusion from Research Articles

Author	Title	Object	Algorithm	Result
(Almansour et al., 2019)	Neural network and support vector machine for the prediction of chronic kidney disease: A comparative study	Measuring the level of accuracy and compare SVM and ANN on the CKD dataset	ANN, SVM	Accuracy for ANN 99.75%, for SVM 97.75%
(Ma et al., 2020)	Detection and diagnosis of chronic kidney disease using deep learning-based heterogeneous modified artificial neural network	HMANN to detect chronic renal disease	Heterogenous Modified Artificial Neural Network (HMANN)	Proposed HMANN algorithm is proposed to detect chronic renal disease
(Ed-daoudy & Maalmi, 2020)	Breast cancer classifications with reduced feature set using association rules and support vector machine	Classifications for breast cancer	Association Rules, SVM	Classification accuracy by SVM model and AR = 98% for 8 attributes and 96.14% for 4 attributes
(López et al., 2018)	Single Nucleotide Polymorphism relevance learning with Random Forest for Type 2 Diabetes risk prediction	SNP relevance with T2D risk prediction	Random Forest	Obtained 14 most relevant SNPs
(B. Yu et al., 2021)	Prediction of protein-protein interactions based on elastic net and deep forest	Method for predicting PPI	Elastic Net, GcForest-PPI	PPI's of a CD9-core network are all predicted successfully, Gc-Forest-PPI able to predict PPIs in a crossover network and able to reveal the biological functions for the Wnt-related pathway, PPIs of the cancer-specific network are also all predicted successfully
(Blair et al., 2014)	On the stability of the Bayenv method in assesing human SNP-environment associations	Bayenv method in scoring association between SNP and environment	He Bayenv (using Bayesian Markov Chain Monte Carlo Approach)	Bayenv produced slightly different results in independent runs of the program
(Liu & Xuan,	Prioritization of Cancer-Related Genomic	SNP association	RWR-SAN	A new method that combines the AMD

2015)	Variants by SNP Association Network	network (SAN)		GWAS dataset with SAN topological information to re-rank the relevance between SNPs and AMD
(Nasarian et al., 2020)	Association between work-related features and coronary artery disease: A heterogeneous hybrid feature selection integrated with balancing approach	Association between work-related feature with coronary artery disease	Decision Tree, Gaussian Naïve Bayes, RF, XGBoost	Proposed feature selection method (SMOTE and XGBoost classifier) achieves accuracy = 81.23%, also tested to other database and obtained high accuracies
(Glazyrin et al., 2020b)	Proteomics-Based Machine Learning Approach as an Alternative to Conventional Biomarkers for Differential Diagnosis of Chronic Kidney Diseases	The usage of full-proteomic to diagnose comparison between chronic kidney disease	K-NN, logistic regression, SVM, decision tree	K-NN algorithm shows high confidence (97.8%) to separate the healthy group from the renal patients in general by proteomics data of plasma
(Antony et al., 2021)	A Comprehensive Unsupervised Framework for Chronic Kidney Disease Prediction	Unsupervised framework to predict CKD	K-Means clustering, DB-Scan, I-Forest, Autoencoder	Integrating feature reduction methods with K-Means Clustering algorithm by achieved accuracy 99% for classifying the clinical data of CKD and non-CKD patients
(Ilyas et al., 2021)	Chronic kidney disease diagnosis using decision tree algorithm	Diagnose CKD stages	RF, J48	J48 algorithm has better result to predicted CKD than RF with an accuracy of 85.5%
(Bradley et al., 2019)	Predicting early risk of chronic kidney disease in cats using routine clinical laboratory tests and machine learning	Predicting the risk of cat with CKD	Architectural neural network	Recurrent Neural Network model with 4 features gain 90.7% of sensitivity, and 98.9% of specificity.
(Zubair Hasan & Zahid Hasan, 2019)	Performance Evaluation of Ensemble-Based Machine Learning Techniques for Prediction of Chronic Kidney Disease	Evaluation in several machine learning techniques classifying CKD patients	Adaptive Boosting, Bootstrap Aggregating, Extra Trees, Gradient Boosting, RF classifier	Ensemble method-based machine learning algorithms to classify for kidney disease and achieves 99% of classification accuracy
(Shang et al., 2021)	Medical records-based chronic kidney disease	Proposed CKD	Combination of rule-based	Algorithm to detect CKD in its early

	phenotype for clinical care and “big data” observational and genetic studies	algorithm	and machine learning method (classification, regression, clustering)	stages by calculating albuminuria and estimated glomerular filtration rate
(Ho et al., 2019)	Machine Learning SNP Based Prediction for Precision Medicine	Machine learning algorithms that have been used for precision medicine	Lasso regression, SVM, NN, decision tree, RF	Overview of polygenic risk scoring and machine learning in complex disease risk prediction
(Yao et al., 2013)	Random Forests approach for identifying additive and epistatic single nucleotide polymorphisms associated with residual feed intake in dairy cattle	Identification SNP additive and epistatic associated with residual feed intake in dairy cows	RF	25 most frequent pairwise SNP interactions were reported as possible epistatic interactions
(Korani et al., 2019)	Machine Learning as an Effective Method for Identifying True Single Nucleotide Polymorphisms in Polyploid Plants	Method to call SNP polyploid species with machine learning	NN, tree bagger	A new 48K array was designed and validated, also created new tool (SNP-MLer) to train models for use in selecting true SNPs from sequence data
(Kucukkal et al., 2014)	Computational and Experimental Approaches to Reveal the Effects of Single Nucleotide Polymorphisms with Respect to Disease Diagnostics	Computational and experimental method to demonstrate the possible effects of genetic differences	SVM	MECP2 and the KDM5C were selected based on this ongoing research
(Hess et al., 2017)	Partitioned learning of deep Boltzmann machines for SNP data	Deep Boltzmann for SNP data	Boltzmann machines (deep learning)	The proposed approach identified 3 SNPs that seem to jointly influence survival in a validation dataset
(De Maturana et al., 2013)	Application of Multi-SNP Approaches Bayesian LASSO and AUC-RF to Detect Main Effects of Inflammatory-Gene Variants Associated	Association between gen inflammation variants and bladder cancer	Bayessian threshold lasso (BTL), AUC-RF	A set of SNP (9 SNP-signature was detected by BTL) in inflammatory genes jointly associated with bladder cancer risk

	with Bladder Cancer Risk			
(Gaudillo et al., 2019)	Machine learning approach to single nucleotide polymorphism-based asthma prediction	Machine learning for predicting the risk of suspecting asthma	RF, RFE, k-NN, SVM	RF-SVM obtained the highest model performance (including accuracy, precision, and sensitivity)
(Ghafouri-Fard et al., 2019)	Application of Single-Nucleotide Polymorphisms in the Diagnosis of Autism Spectrum Disorders: A Preliminary Study with Artificial Neural Networks	Predicting ASD status	ANN	Proposed model with ANN-based for separate ASD status (healthy and adequate power) and gave 73.67% of accuracy, 82.75% of sensitivity, and 63.95% of specificity
(Ghafouri-Fard et al., 2020)	Application of Artificial Neural Network for Prediction of Risk of Multiple Sclerosis Based on Single Nucleotide Polymorphism Genotypes	Predicting risk of multiple sclerosis	ANN	ANN model for prediction the risk of MS based on genomic data. Accuracy = 64.73% Sensitivity = 64.95% Specificity = 64.49%
(F. Yu et al., 2014)	Differentially-Private Logistic Regression for Detecting Multiple-SNP Association in GWAS Databases	Detection the association of multi-SNP on GWAS database	Elastic net regression	Method that is applicable to GWAS datasets and enables to perform data analysis that preserves privacy and utility
(Burnard et al., 2022)	Capturing SNP Association across the NK Receptor and HLA Gene Regions in Multiple Sclerosis by Targeted Penalized Regression Models	SNP association across the NK receptor and HLA gene regions in multiple sclerosis	Penalized regression	Alternate approach to analyzing GWAS data
(Xiao et al., 2019)	Comparison and development of machine learning tools in the prediction of chronic kidney disease progression	Machine learning tools in predicting CKD progress	logistic regression, elastic net, lasso, ridge, svm, random forest, XGBoost, neural network, k-nearest neighbor	The model with the highest sensitivity was Elastic Net (0.85), while XGBoost showed the highest specificity (0.83). In the effect size analyses, we identified that ALB, Scr, TG, LDL and EGFR had important impacts on

				the predictability of the models
(Ing et al., 2017)	A novel machine learning approach reveals latent vascular phenotypes predictive of renal cancer outcome	Vascular latent phenotype in prediction of kidney cancer outcome	SVM, RF, GLMNET	Identified a 14 gene expression signature related to the 9VF's
(Dessie et al., 2021)	A novel miRNA-based classification model of risks and stages for clear cell renal cell carcinoma patients	miRNA classification about risks and levels for clear cell renal cell carcinoma patients	Lasso, elastic net, logistic regression, rf, svm, averaging neural network, NB, k-NN	Identified a four-miRNA signature associated with survival, developed a combined ML algorithms to identify six miRNA signatures for cancer staging prediction
(Hayes et al., 2021)	Prediction of individuals at high risk of chronic kidney disease during treatment with lithium for bipolar disorder	Predicting of CKD patients who use lithium for bipolar disorder treatment	Elastic Net	A model to predict, at lithium initiation, individuals at high risk for poor renal function trajectory using serial eGFR measurement
(Ebiaredoh-Mienye et al., 2022)	A Machine Learning Method with Filter-Based Feature Selection for Improved Prediction of Chronic Kidney Disease	Feature selection with filter-based to predict CKD	AdaBoost	Proposed approach with AdaBoost classifier based to detect CKD with 99.8% of accuracy, 100% of sensitivity, and 99.8% of specificity
(Senan et al., 2021)	Diagnosis of Chronic Kidney Disease Using Effective Classification Algorithms and Recursive Feature Elimination Techniques	Diagnose system to detect CKD	RFE, SVM, k-NN, decision tree, RF	RF algorithm achieved 100% of accuracy, precision, recall, and F1-score
(N et al., 2019)	An ensemble multi-model technique for predicting chronic kidney disease	An ensemble multi-model technique to predict CKD	SVM, Decision tree, C4.5, PSO-MLP	ABC4.5 model to detect CKD
(Dritsas & Trigka, 2022)	Machine Learning Techniques for Chronic Kidney Disease Risk Prediction	Machine learning techniques to predict CKD risk	K-NN, SVM, LR, SGD, ANN	Tools for predicting CKD occurrence called Rotation Forest (RotF), with AUC 100%, Precision, Recall, F-Measure, and accuracy reached 99,2%

(Akben, 2018)	Early Stage Chronic Kidney Disease Diagnosis by Applying Data Mining Methods to Urinalysis, Blood Analysis and Disease History	Diagnosing early-stage CKD with attention to urinalysis, blood counts, and history	k-means clustering, k-NN, SVM, naïve bayes	A new automatic early-stage CKD diagnosis method
(da Silveira et al., 2022)	Exploring Early Prediction of Chronic Kidney Disease Using Machine Learning Algorithms for Small and Imbalanced Datasets	Early prediction of CKD with small and imbalanced dataset	DT, RF, AdaBoost, k-NN	DT model presented the highest accuracy score = 98.99%, with purpose to assist the early prediction of CKD that related to imbalanced and limited-size datasets
(Kao et al., 2022)	Associations between Sex and Risk Factors for Predicting Chronic Kidney Disease	Gender relationship with risk factors predicting CKD	SVM, LDA, LR, C4.5 DT, CART, RF, C5.0 DT	A prediction model to support diagnosis and treatment for early stage of CKD. The model achieving acceptable prediction accuracy
(Lee et al., 2022)	Development and Validation of an Insulin Resistance Model for a Population with Chronic Kidney Disease Using a Machine Learning Approach	Insulin resistance CKD model	RF, XGBoost, logistic regression, DNN	In the validation group, RF had a higher accuracy (0.77), specificity (0.81), PPV (0.77), and NPV (0.77). In the test group, XGboost had a higher AUC of ROC (0.78). In addition, XGBoost also had a higher accuracy (0.7) and NPV (0.71). RF had a higher accuracy (0.7), specificity (0.78), and PPV (0.7). RF algorithm had the best AUC of ROC and the best SHAP value differentiation

From the reviews conducted, it can be concluded that classification is the most widely used machine learning approach in precision medicine research, such as random forest algorithms and support vector machines are two of the most commonly used algorithms for classification. In this systematic literature review, the researchers are offered helps in developing or expanding a model in medical field, especially in precision medicine, with using machine learning technique approach. Unfortunately, most of the reviewed articles in this research used supervised machine learning, so that it will need further related research using unsupervised machine learning to gain new and more information.

Most of the articles that have been reviewed aim to develop or build a model with machine learning approach to predict a certain complex disease. To assure the success of the algorithm or machine learning approach model, several articles have implemented several measurements such as the level of accuracy, level of specificity, AUC, precision, F-measure, recall, and sensitivity. Still, there has not been a clear protocol for emerging machine learning in healthcare (Sadat Mosavi and Filipe Santos, 2021). Besides that, ethical consideration will arise from the emerging, complex and cross-disciplinary technology that can affect many aspects of healthcare. These ethical considerations can be solved by promoting benefit and clarifying the value of ML (Frank, Matthew G. Annis, Watkins, 2019).

5. Conclusion

This study reviews research articles that discuss about machine learning approaches for precision medicine with the aim of discovering the journal publisher, search keywords, machine learning algorithms and types that are widely used in research related to machine learning approaches for precision medicine to journals that have been published between 2013 – 2022.

Based on the results of this review conducted using systematic literature review method, it is known that the most widely used machine learning algorithm is random forest, followed by support vector machine, decision tree, k-nearest neighbor, elastic net, and artificial neural network. Meanwhile machine learning, chronic kidney disease, SNP, and prediction are the most used search keywords to find articles related to machine learning approaches for precision medicine.

Acknowledgment

The authors would like to thank Universitas Multimedia Nusantara for the support of this research work.

References

- Akben, S. B. (2018). Early Stage Chronic Kidney Disease Diagnosis by Applying Data Mining Methods to Urinalysis, Blood Analysis and Disease History. *Irbm*, 39(5), 353–358. <https://doi.org/10.1016/j.irbm.2018.09.004>
- Almansour, N. A., Syed, H. F., Khayat, N. R., Altheeb, R. K., Juri, R. E., Alhiyafi, J., Alrashed, S., & Olatunji, S. O. (2019). Neural network and support vector machine for the prediction of chronic kidney disease: A comparative study. *Computers in Biology and Medicine*, 109, 101–111. <https://doi.org/10.1016/j.combiomed.2019.04.017>
- Antony, L., Azam, S., Ignatious, E., Quadir, R., Beeravolu, A. R., Jonkman, M., & De Boer, F. (2021). A Comprehensive Unsupervised Framework for Chronic Kidney Disease Prediction. *IEEE Access*, 9, 126481–126501. <https://doi.org/10.1109/ACCESS.2021.3109168>
- Blair, L. M., Granka, J. M., & Feldman, M. W. (2014). On the stability of the Bayenv method in assessing human SNP-environment associations. *Human Genomics*, 8(1). <https://doi.org/10.1186/1479-7364-8-1>
- Bradley, R., Tagkopoulos, I., Kim, M., Kokkinos, Y., Panagiotakos, T., Kennedy, J., De Meyer, G., Watson, P., & Elliott, J. (2019). Predicting early risk of chronic kidney disease in cats using routine clinical laboratory tests and machine learning. *Journal of Veterinary Internal Medicine*, 33(6), 2644–2656. <https://doi.org/10.1111/jvim.15623>

- Burnard, S. M., Lea, R. A., Benton, M., Eccles, D., Kennedy, D. W., Lechner-Scott, J., & Scott, R. J. (2022). Capturing SNP Association across the NK Receptor and HLA Gene Regions in Multiple Sclerosis by Targeted Penalised Regression Models. *Genes*, 13(1). <https://doi.org/10.3390/genes13010087>
- Carvalho, T. P., Soares, F. A. A. M. N., Vita, R., Francisco, R. da P., Basto, J. P., & Alcalá, S. G. S. (2019). A systematic literature review of machine learning methods applied to predictive maintenance. *Computers and Industrial Engineering*, 137(August), 106024. <https://doi.org/10.1016/j.cie.2019.106024>
- Clish, C. B. (2015). Metabolomics: an emerging but powerful tool for precision medicine. *Molecular Case Studies*, 1(1), a000588. <https://doi.org/10.1101/mcs.a000588>
- da Silveira, A. C. M., Sobrinho, Á., da Silva, L. D., Costa, E. de B., Pinheiro, M. E., & Perkusich, A. (2022). Exploring Early Prediction of Chronic Kidney Disease Using Machine Learning Algorithms for Small and Imbalanced Datasets. *Applied Sciences (Switzerland)*, 12(7), 3673. <https://doi.org/10.3390/app12073673>
- De Maturana, E. L., Ye, Y., Calle, M. L., Rothman, N., Urrea, V., Kogevinas, M., Petrus, S., Chanock, S. J., Tardón, A., García-Closas, M., González-Neira, A., Vellalta, G., Carrato, A., Navarro, A., Lorente-Galdós, B., Silverman, D. T., Real, F. X., Wu, X., & Malats, N. (2013). Application of multi-SNP approaches Bayesian LASSO and AUC-RF to detect main effects of inflammatory-gene variants associated with bladder cancer risk. *PLoS ONE*, 8(12). <https://doi.org/10.1371/journal.pone.0083745>
- Dessie, E. Y., Tsai, J. J. P., Chang, J. G., & Ng, K. L. (2021). A novel miRNA-based classification model of risks and stages for clear cell renal cell carcinoma patients. *BMC Bioinformatics*, 22, 1–15. <https://doi.org/10.1186/s12859-021-04189-2>
- Dritsas, E., & Trigka, M. (2022). Machine Learning Techniques for Chronic Kidney Disease Risk Prediction. *Big Data and Cognitive Computing*, 6(3), 98. <https://doi.org/10.3390/bdcc6030098>
- Ebiaredoh-Mienye, S. A., Swart, T. G., Esenogho, E., & Mienye, I. D. (2022). A Machine Learning Method with Filter-Based Feature Selection for Improved Prediction of Chronic Kidney Disease. *Bioengineering*, 9(8), 350. <https://doi.org/10.3390/bioengineering9080350>
- Ed-daoudy, A., & Maalmi, K. (2020). Breast cancer classification with reduced feature set using association rules and support vector machine. *Network Modeling Analysis in Health Informatics and Bioinformatics*, 9(1). <https://doi.org/10.1007/s13721-020-00237-8>
- Gard, J. J., Cha, Y. M., & Friedman, P. A. (2013). Leadless pacing and defibrillation systems. *Cardiac Electrophysiology Clinics*, 5(3), 327–335. <https://doi.org/10.1016/j.ccep.2013.05.009>
- Gaudillo, J., Joseph Russell Rodriguez, J., Nazareno, A., Rigi Baltazar, L., Vilela, J., Bulalacao, R., Domingo, M., & Albia, J. (2019). Machine learning approach to single nucleotide polymorphism-based asthma prediction. *PLoS ONE*, 14(12), 1–12. <https://doi.org/10.1371/journal.pone.0225574>
- Ghafouri-Fard, S., Taheri, M., Omrani, M. D., Daaee, A., & Mohammad-Rahimi, H. (2020). Application of Artificial Neural Network for Prediction of Risk of Multiple Sclerosis Based on Single Nucleotide Polymorphism Genotypes. *Journal of Molecular Neuroscience*, 70(7), 1081–1087. <https://doi.org/10.1007/s12031-020-01514-x>
- Ghafouri-Fard, S., Taheri, M., Omrani, M. D., Daaee, A., Mohammad-Rahimi, H., & Kazazi, H. (2019). Application of Single-Nucleotide Polymorphisms in the Diagnosis of Autism Spectrum Disorders: A Preliminary Study with Artificial Neural Networks. *Journal of Molecular Neuroscience*, 68(4), 515–521. <https://doi.org/10.1007/s12031-019-01311-1>
- Glazyrin, Y. E., Veprintsev, D. V., Ler, I. A., Rossovskaya, M. L., Varygina, S. A., Glizer, S. L., Zamay, T. N., Petrova, M. M., Minic, Z., Berezovski, M. V., & Kichkailo, A. S. (2020). Proteomics-based machine learning approach as an alternative to conventional biomarkers for differential diagnosis of chronic kidney diseases. *International Journal of Molecular Sciences*, 21(13), 1–11. <https://doi.org/10.3390/ijms21134802>

- Hayes, J. F., Osborn, D. P. J., Francis, E., Ambler, G., Tomlinson, L. A., Boman, M., Wong, I. C. K., Geddes, J. R., Dalman, C., & Lewis, G. (2021). Prediction of individuals at high risk of chronic kidney disease during treatment with lithium for bipolar disorder. *BMC Medicine*, *19*(1), 1–16. <https://doi.org/10.1186/s12916-021-01964-z>
- Hess, M., Lenz, S., Blätte, T. J., Bullinger, L., & Binder, H. (2017). Partitioned learning of deep Boltzmann machines for SNP data. *Bioinformatics*, *33*(20), 3173–3180. <https://doi.org/10.1093/bioinformatics/btx408>
- Ho, D. S. W., Schierding, W., Wake, M., Saffery, R., & O’Sullivan, J. (2019). Machine learning SNP based prediction for precision medicine. *Frontiers in Genetics*, *10*(MAR). <https://doi.org/10.3389/fgene.2019.00267>
- Ilyas, H., Ali, S., Ponum, M., Hasan, O., Mahmood, M. T., Iftikhar, M., & Malik, M. H. (2021). Chronic kidney disease diagnosis using decision tree algorithms. *BMC Nephrology*, *22*(1), 273. <https://doi.org/10.1186/s12882-021-02474-z>
- Ing, N., Huang, F., Conley, A., You, S., Ma, Z., Klimov, S., Ohe, C., Yuan, X., Amin, M. B., Figlin, R., Gertych, A., & Knudsen, B. S. (2017). A novel machine learning approach reveals latent vascular phenotypes predictive of renal cancer outcome. *Scientific Reports*, *7*(1), 1–10. <https://doi.org/10.1038/s41598-017-13196-4>
- Kao, H. Y., Chang, C. C., Chang, C. F., Chen, Y. C., Cheewakriangkrai, C., & Tu, Y. L. (2022). Associations between Sex and Risk Factors for Predicting Chronic Kidney Disease. *International Journal of Environmental Research and Public Health*, *19*(3), 1219. <https://doi.org/10.3390/ijerph19031219>
- Khurana Hershey, G. K., Sherenian, M. G., & Mersha, T. B. (2022). Precision Medicine. *Allergy Essentials*, 25–39. <https://doi.org/10.1016/B978-0-323-80912-2.00002-0>
- Klonoff, D. C. (2015). Precision medicine for managing diabetes. *Journal of Diabetes Science and Technology*, *9*(1), 3–7. <https://doi.org/10.1177/1932296814563643>
- Korani, W., Clevenger, J. P., Chu, Y., & Ozias-Akins, P. (2019). Machine Learning as an Effective Method for Identifying True Single Nucleotide Polymorphisms in Polyploid Plants. *The Plant Genome*, *12*(1), 180023. <https://doi.org/10.3835/plantgenome2018.05.0023>
- Kucukkal, T. G., Yang, Y., Chapman, S. C., Cao, W., & Alexov, E. (2014). Computational and experimental approaches to reveal the effects of single nucleotide polymorphisms with respect to disease diagnostics. In *International Journal of Molecular Sciences* (Vol. 15, Issue 6). <https://doi.org/10.3390/ijms15069670>
- Lee, C. L., Liu, W. J., & Tsai, S. F. (2022). Development and Validation of an Insulin Resistance Model for a Population with Chronic Kidney Disease Using a Machine Learning Approach. *Nutrients*, *14*(14), 1–18. <https://doi.org/10.3390/nu14142832>
- Liu, C., & Xuan, Z. (2015). Prioritization of cancer-related genomic variants by SNP association network. *Cancer Informatics*, *14*, 57–70. <https://doi.org/10.4137/CIN.S17288>
- López, B., Torrent-Fontbona, F., Viñas, R., & Fernández-Real, J. M. (2018). Single Nucleotide Polymorphism relevance learning with Random Forests for Type 2 diabetes risk prediction. *Artificial Intelligence in Medicine*, *85*, 43–49. <https://doi.org/10.1016/j.artmed.2017.09.005>
- Ma, F., Sun, T., Liu, L., & Jing, H. (2020). Detection and diagnosis of chronic kidney disease using deep learning-based heterogeneous modified artificial neural network. *Future Generation Computer Systems*, *111*, 17–26. <https://doi.org/10.1016/j.future.2020.04.036>
- Masic, I., Miokovic, M., & Muhamedagic, B. (2008). Evidence Based Medicine - New Approaches and Challenges. *Acta Informatica Medica*, *16*(4), 219. <https://doi.org/10.5455/aim.2008.16.219-225>

- N, K. K., Tulasi, R. L., & D, V. (2019). An ensemble multi-model technique for predicting chronic kidney disease. *International Journal of Electrical and Computer Engineering (IJECE)*, 9(2), 1321. <https://doi.org/10.11591/ijece.v9i2.pp1321-1326>
- Nasarian, E., Abdar, M., Fahami, M. A., Alizadehsani, R., Hussain, S., Basiri, M. E., Zomorodi-Moghadam, M., Zhou, X., Pławiak, P., Acharya, U. R., Tan, R. S., & Sarrafzadegan, N. (2020). Association between work-related features and coronary artery disease: A heterogeneous hybrid feature selection integrated with balancing approach. *Pattern Recognition Letters*, 133, 33–40. <https://doi.org/10.1016/j.patrec.2020.02.010>
- Senan, E. M., Al-Adhaileh, M. H., Alsaade, F. W., Aldhyani, T. H. H., Alqarni, A. A., Alsharif, N., Uddin, M. I., Alahmadi, A. H., Jadhav, M. E., & Alzahrani, M. Y. (2021). Diagnosis of Chronic Kidney Disease Using Effective Classification Algorithms and Recursive Feature Elimination Techniques. *Journal of Healthcare Engineering*, 2021, 1–10. <https://doi.org/10.1155/2021/1004767>
- Shang, N., Khan, A., Polubriaginof, F., Zaroni, F., Mehl, K., Fasel, D., Drawz, P. E., Carrol, R. J., Denny, J. C., Hathcock, M. A., Arruda-Olson, A. M., Peissig, P. L., Dart, R. A., Brilliant, M. H., Larson, E. B., Carrell, D. S., Pendergrass, S., Verma, S. S., Ritchie, M. D., ... Kiryluk, K. (2021). Medical records-based chronic kidney disease phenotype for clinical care and “big data” observational and genetic studies. *Npj Digital Medicine*, 4(1). <https://doi.org/10.1038/s41746-021-00428-1>
- Xiao, J., Ding, R., Xu, X., Guan, H., Feng, X., Sun, T., Zhu, S., & Ye, Z. (2019). Comparison and development of machine learning tools in the prediction of chronic kidney disease progression. *Journal of Translational Medicine*, 17(1), 1–13. <https://doi.org/10.1186/s12967-019-1860-0>
- Yao, C., Spurlock, D. M., Armentano, L. E., Page, C. D., VandeHaar, M. J., Bickhart, D. M., & Weigel, K. A. (2013). Random Forests approach for identifying additive and epistatic single nucleotide polymorphisms associated with residual feed intake in dairy cattle. *Journal of Dairy Science*, 96(10), 6716–6729. <https://doi.org/10.3168/jds.2012-6237>
- Yu, B., Chen, C., Wang, X., Yu, Z., Ma, A., & Liu, B. (2021). Prediction of protein–protein interactions based on elastic net and deep forest. *Expert Systems with Applications*, 176. <https://doi.org/10.1016/j.eswa.2021.114876>
- Yu, F., Rybar, M., Uhler, C., & Fienberg, S. E. (2014). Differentially-private logistic regression for detecting multiple-SNP association in GWAS databases. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8744, 170–184. https://doi.org/10.1007/978-3-319-11257-2_14
- Zubair Hasan, K. M., & Zahid Hasan, M. (2019). Performance Evaluation of Ensemble-Based Machine Learning Techniques for Prediction of Chronic Kidney Disease. In *Advances in Intelligent Systems and Computing* (Vol. 882). Springer Singapore. https://doi.org/10.1007/978-981-13-5953-8_34