Algorithmic models with artificial intelligence for disease diagnosis: A systematic literature review

ISBN: 978-628-96613-1-6. ISSN: 2414-6390. Digital Object Identifier: https://dx.doi.org/10.18687/LACCEI2025.1.1.548

Fernandez-Zuloeta Alvaro[1]; Morales-Vega Jose[2]; Soria-Quijaite Juan[3]; Yapo-Caceres Carlos[4]
[1,2,3,4]Universidad Tecnológica del Perú, *U21201804@utp.edu.pe, U21101125@utp.edu.pe, c20723@utp.edu.pe, c18236@utp.edu.pe*

Abstract: The integration of artificial intelligence (AI) in disease diagnosis is transforming the field of medicine, offering precise and efficient tools to detect critical health conditions. However, the diversity of algorithms and their applications raises questions about which of these are the most effective. Therefore, this systematic review aims to identify the most accurate AI models for detecting various diseases. A systematic review was conducted considering Scopus and Web of Science databases as main information sources, analyzing 416 studies that used AI algorithms within the medical field. Additionally, rigorous inclusion and exclusion criteria were applied, screening 26 articles that prioritize quantitative results relevant to clinical diagnosis. The most prominent model is Random Forest, with a frequency of use in 12 investigations and an average accuracy of 0.91. Likewise, the XGBoost, CNN, and SVM models were used in 9 investigations each and obtained accuracy results of 0.87, 0.88, and 0.96 respectively. These performances were particularly notable in applications related to dermatological, cardiological, and oncological diseases. The results position Random Forest as an efficient tool for medical diagnosis, although its practical implementation faces some technological and budgetary challenges. It is recommended to explore hybrid methodologies that combine advanced algorithms with more traditional approaches and conduct longitudinal studies to evaluate their impact in different clinical settings.

Keywords—Artificial Intelligence, Algorithm, Machine Learning, Deep Learning, Disease

I. INTRODUCTION

Machine learning (ML) algorithmic models are types of artificial intelligence that allow a computer program to obtain information from errors, analyze data, recognize patterns, and produce informed judgments with minimal human intervention [1]. The use of this technology has revolutionized predictive work across many sectors, with healthcare being one of the most notable, where the ability to diagnose diseases with the support of computer systems has been enhanced.

The impact of AI in the medical sector is largely recognized thanks to the wide diversity of scientific literature related to its use in the diagnosis of various diseases. Currently, a total of 3,508 articles have been identified globally on the application of AI in rheumatic diseases, with the United States and China being the nations that contribute the highest number of publications, followed by the United Kingdom, Germany, and South Korea. [2]. On the other hand, in countries like Mexico, ML techniques such as semi-supervised learning have been applied for the segmentation and estimation of lung lesions caused by COVID-19, where local patient data was used to demonstrate the viability of these technologies in medical diagnosis. [3]. Likewise, in ophthalmological medical centers in Lima, Peru, deep learning (DL) was used as a tool for medical diagnoses. DL belongs to a subgroup of algorithms within ML and is based on the use of neural networks to make predictions. Among the different types of DL techniques, the convolutional neural network (CNN) algorithm achieved highly effective results, exceeding 85% accuracy in the diagnosis of eye diseases. [4]. AI applications in the healthcare sector continue to increase, and the use of different ML or DL techniques is becoming increasingly popular in medical practice across different parts of the world.

Although today, these algorithmic models are not capable of replacing a doctor's work when identifying a disease [5], paradigms such as the use of support vector machines to predict patient response to cancer treatments [6] show that these models serve as important tools against misdiagnosis. Similarly, models based on DL and neural networks are another type of artificial intelligence technology that face significant challenges that keep them far from medical practice, such as the need to have a large amount of training data for effective operation [7]; however, the successful use of convolutional neural networks (CNN) in classification and detection processes related to medical images [8], have demonstrated that they could be tools of great potential for clinical treatments, showing accuracy values of up to 98.0% for the classification of breast lesions into benign and malignant [9].

This Systematic Literature Review on the use of artificial intelligence for disease diagnosis is important for the software engineering field not only because AI is widely used and has shown high growth in the last decade [10]; but also because it provides a detailed description of the different ML and DL technologies used in medical diagnostic processes, such as neural networks, SVM, logistic regression, decision trees, and random forests in cases of cardiac problems [11]. In this sense, recent studies demonstrate the success of DL in various medical applications, affirming the need to expand scientific literature to validate the use of these new approaches in clinical practice.

II. METHODOLOGY

The structure of this document follows the Systematic Literature Review (SLR) format, aiming to compile different sources related to the topic in question and group them together so that they collectively obtain a new meaning that will add value to scientific literature.

The PICOC method shown in Table 1 was used to define the scope of this research. This strategy allowed us to identify concepts and key words in different dimensions to finally find a research question.

Table 1. Description of the PICOC method

P	Population/Problem	Medical diagnosis of diseases
I	Intervention	Algorithmic models of artificial intelligence
C	Comparison	Machine Learning and Deep Learning Algorithmic Models
O	Outcomes	Accuracy Level
C	Context	In Healthcare Centers

A. RESEARCH QUESTION

Through the PICOC method and clearly defined concepts, the main question of the SLR seeks to answer: What algorithmic models of artificial intelligence are used to determine the accuracy level in medical diagnosis of diseases in healthcare centers?

B. REVIEW SUB-QUESTIONS

RQ1: What type of prediction problem is the most recurrent in the use of artificial intelligence algorithmic models for disease diagnosis?

RQ2: What is the accuracy level of artificial intelligence algorithmic models in disease diagnosis?

RQ3: In what types of healthcare centers have machine learning and deep learning models been applied for disease diagnosis?

C. INCLUSION AND EXCLUSION CRITERIA

In order to compile articles that contribute to answering the research question, inclusion and exclusion criteria were declared that served as a guide for publication search. The inclusion criteria considered that all empirical articles collected must be focused on the use of artificial intelligence algorithmic models. Likewise, they must compare performance metrics of different ML models, reporting their accuracy levels with a quantitative approach. The SLR only used studies conducted or applied in healthcare centers and related to the computer science field. On the other hand, the exclusion criteria determined all documents published before 2024 or in a language other than English or Spanish. Similarly, research that uses artificial intelligence for other aspects of medical care that are not disease diagnosis or related to COVID-19 diseases will be discarded.

D. PUBLICATION SOURCES

For the compilation of the different publications in this study, the Scopus and Web of Science databases were used as sources, considered two of the most recognized in scientific research environments.

1) SEARCH EQUATION FOR SCOPUS:

The search equation for the Scopus database is represented by logical connectors and is:

(TITLE-ABS-KEY ("medical diagnosis" OR "analysis" OR "disease diagnosis" OR "clinical diagnosis" OR "diagnosis software" OR "pathological diagnosis" OR "computer-assisted diagnosis" OR "computer-aided diagnosis" OR "medical radiological diagnosis") AND TITLE-ABS-KEY ("Algorithm" OR "artificial intelligence") AND TITLE-ABS-KEY ("machine learning" OR "supervised learning" OR "unsupervised learning" OR "self-adapting computer" OR "ML system" OR "deep learning" OR "neural network" OR "neural net" OR "recurrent neural network" OR "RNN" OR "multilayer neural network" OR "convolutional neural network" OR "CNN" OR "Dictionary learning" OR "Cognitive systems" OR "Data augmentation" OR "Deep architecture" OR "Image augmentation" OR "Linear discriminant analysis" OR "Long short term memory" OR "Naive Bayes methods" OR "Photorealistic images" OR "Predictive analytics" OR "Radiomics" OR "Boosting" OR "Deep reinforcement" OR "Diffusion models" OR "Ensemble learning" OR "Hyperparameter" OR "Random forests" OR "Reinforcement learning" OR "Relevance vector machines" OR "Statistical learning" OR "Hierarchical learning" OR "Graph neural networks" OR "Image classification" OR "Image segmentation" OR "Learning systems" OR "Batch normalization") AND TITLE-ABS-KEY ("result" OR "performance" OR "metrics" OR "accuracy" OR "precision" OR "recall" OR "F1 score" OR "ROC curve" OR "AUC" OR "Confusion matrix" OR "Mean Squared Error" OR "MSE" OR "Root Mean Squared Error" OR "RMSE" OR "Mean Absolute Error" OR "MAE" OR "specificity" OR "sensitivity" OR "true positive rate" OR "false positive rate" OR "R-squared" OR "R^2") AND TITLE-ABS-KEY ("clinic" OR "hospital" OR "Medical centers" OR "Health facilities" OR "Primary care centers")) AND PUBYEAR > 2022 AND PUBYEAR < 2026 AND (LIMIT-TO (SUBJAREA , "COMP")) AND (LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOCTYPE , "cp") OR LIMIT-TO (DOCTYPE , "ch")) AND (LIMIT-TO (OA , "all")) AND (LIMIT-TO (LANGUAGE , "English"))

2) SEARCH EQUATION FOR WEB OF SCIENCE:

The search equation for the Web of Science database is represented by logical connectors and is:

"medical diagnosis" OR "analysis" OR "disease diagnosis" OR "clinical diagnosis" OR "diagnosis software" OR "pathological diagnosis" OR "computer-assisted diagnosis" OR "computer-aided diagnosis" OR "medical radiological diagnosis" (All Fields) and "Algorithm" OR "artificial intelligence" (All Fields) and "machine learning" OR "supervised learning" OR "unsupervised learning" OR "self-adapting computer" OR "ML system" OR "deep learning" OR "neural network" OR "neural net" OR "recurrent neural network" OR "RNN" OR "multilayer neural network" OR "convolutional neural network" OR "CNN" OR "Dictionary learning" OR "Cognitive systems" OR "Data augmentation" OR "Deep architecture" OR "Image augmentation" OR "Linear discriminant analysis" OR "Long short term memory" OR "Naive Bayes methods" OR "Photorealistic images" OR "Predictive analytics" OR "Radiomics" OR "Boosting" OR "Deep reinforcement" OR "Diffusion models" OR "Ensemble learning" OR "Hyperparameter" OR "Random forests" OR "Reinforcement learning" OR "Relevance vector machines" OR "Statistical learning" OR "Hierarchical learning" OR "Graph neural networks" OR "Image classification" OR "Image segmentation" OR "Learning systems" OR "Batch normalization" (All Fields) and "result" OR "performance"

OR "metrics" OR "accuracy" OR "precision" OR "recall" OR "F1 score" OR "ROC curve" OR "AUC" OR "Confusion matrix" OR "Mean Squared Error" OR "MSE" OR "Root Mean Squared Error" OR "RMSE" OR "Mean Absolute Error" OR "MAE" OR "specificity" OR "sensitivity" OR "true positive rate" OR "false positive rate" OR "R-squared" OR "R^2" (All Fields) and "clinic" OR "hospital" OR "Medical centers" OR "Health facilities" OR "Primary care centers" (All Fields) and Open Access and 2025 or 2024 or 2023 (Publication Years) and Article (Document Types) and English (Languages) and Computer Science (Research Areas).

E. THE PRISMA STATEMENT

After compiling a total of 465 publications based on the search equations in the Scopus and Web of Science databases, the PRISMA statement diagram shown in *Figure 1* was used as a guide to filter out articles outside the scope of the research. This method consists of three subtitles: Identification, Screening, and Inclusion.

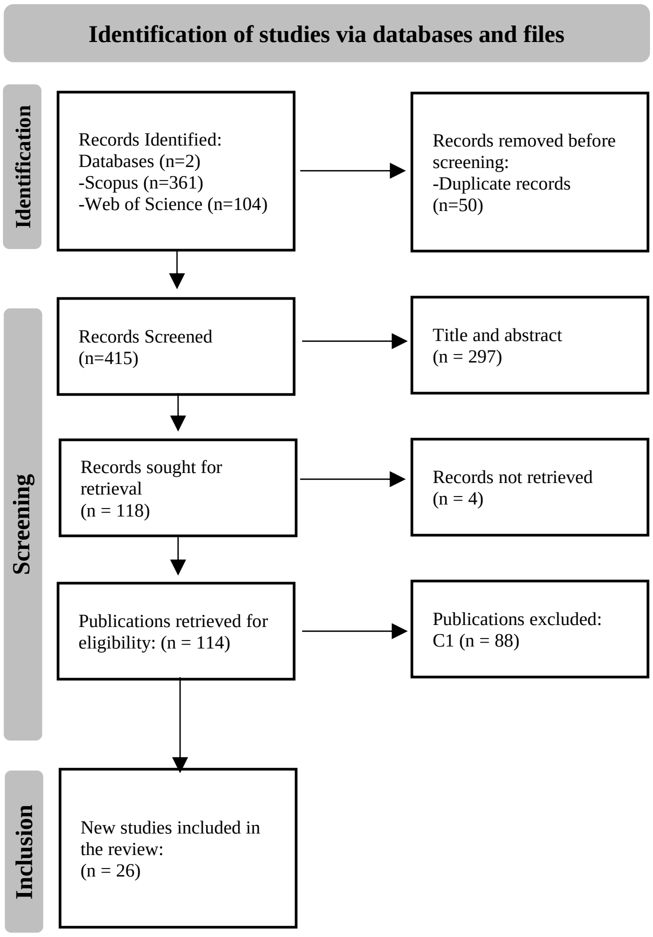

Figure 1: Prisma statement diagram

Identification is a pre-screening process that consisted of eliminating 50 duplicates found after consolidating both databases. After removing duplicate publications, screening continued. Within this process, 297 publications with titles and abstracts not strictly related to the research topic were eliminated. Likewise, 4 articles were left out because they could not be retrieved. Finally, through inclusion and exclusion criteria, another 88 articles were removed. After identification and screening, the inclusion of the PRISMA statement highlights the number of publications selected for the research, reaching a total of 26.

III. RESULTS

A. EXPLORATORY ANALYSIS OF SOURCES

The source selection process for this SLR resulted in 26 articles out of 2024 shown in Table 2. The countries where these publications were produced are China, India, USA, Italy, Oman, South Korea, Bangladesh, Saudi Arabia, Turkey, Belgium, Mexico and Taiwan.

Table 2. Screened articles

REFERENCE	H-INDEX	COUNTRY
Lee et al. [12]	23	United States
Gulamali et al. [13]	7	United States
Parola et al. [14]	2	Italy
Dubbioso et al. [15]	21	Italy
Zhang et al. [16]	69	China
Hong et al. [17]	29	South Korea
Zhu et al. [18]	1	China
Wu et al. [19]	15	China
Nillmani et al. [20]	4	India
Moni et al. [21]	59	Bangladesh
Abdelhafez et al. [22]	1	Saudi Arabi
Qasem et al. [23]	20	United States
Zhang et al. [24]	2	China
Panigrahi et al. [25]	14	India
Cheng et al. [26]	6	China
Fırat et al. [27]	8	Ankara, Turkey
Gudigar et al. [28]	26	India
Alaraimi et al. [29]	1	Oman
Meena et al. [30]	10	India
Karuppasamy et al. [31]	3	Oman
Mayrose et al. [32]	1	India
Bhagubai et al. [33]	4	Belgium
Cabrera Gaytán et al. [34]	5	Mexico
Li et al. [35]	1	China
Chang et al. [36]	22	Taiwan
Hu et al. [37]	19	China

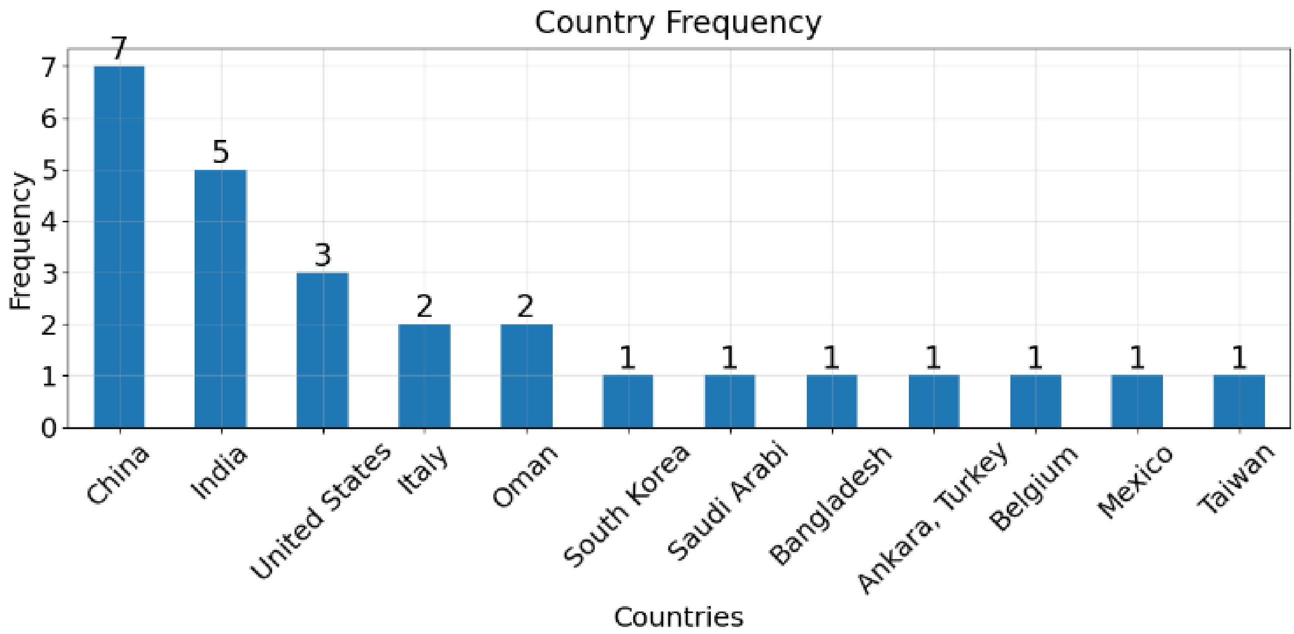

Figure 2: Frequency of countries per item

Figure 2 shows that the countries with the most publications are China and India with seven and five publications respectively, followed by the United States with three.

The number of references used by each article is presented in *Figure 3*, where each bar color corresponds to a specific country. It is observed that the highest number of references used belongs to Lee et al. from the United States (76), followed by Gudigar et al. from India (73), and in third place Zhang et al. from China (68).

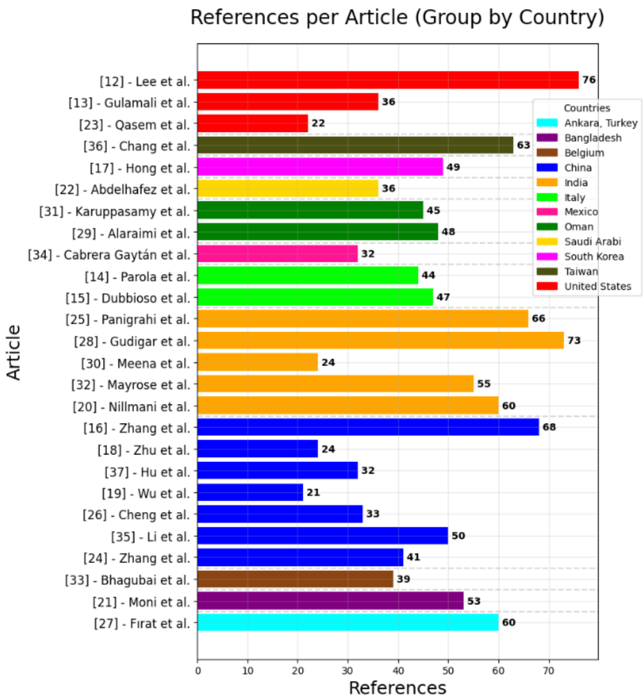

Figure 3: References by article and country

The median impact factors of the articles used in this systematic literature review are between 4 and 5. Likewise, *Figure 4* shows that 50% of publications have a Cite Score between 4 and 10, an impact factor between 3 and 6, and the highest proportion of authors' h-index is between 3 and 20.

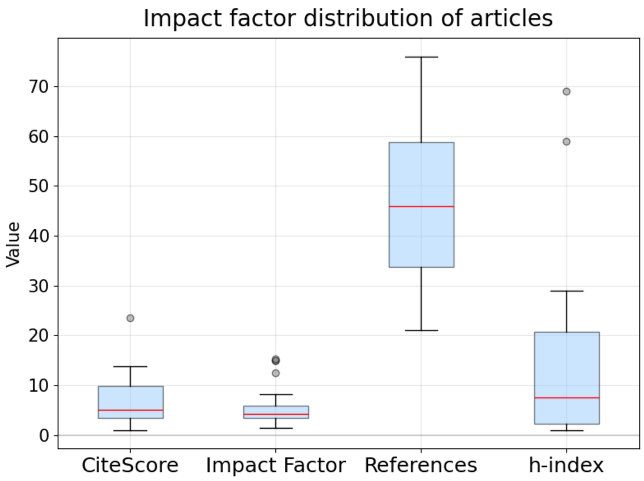

Figure 4: Distribution of article impact factors

Figure 5 shows the set of most recurring keywords in the collection of publications used for this SLR, highlighting the terms machine learning, deep learning, human, article, artificial intelligence, major clinical study, and diagnosis.

Figure 5: Keywords word cloud

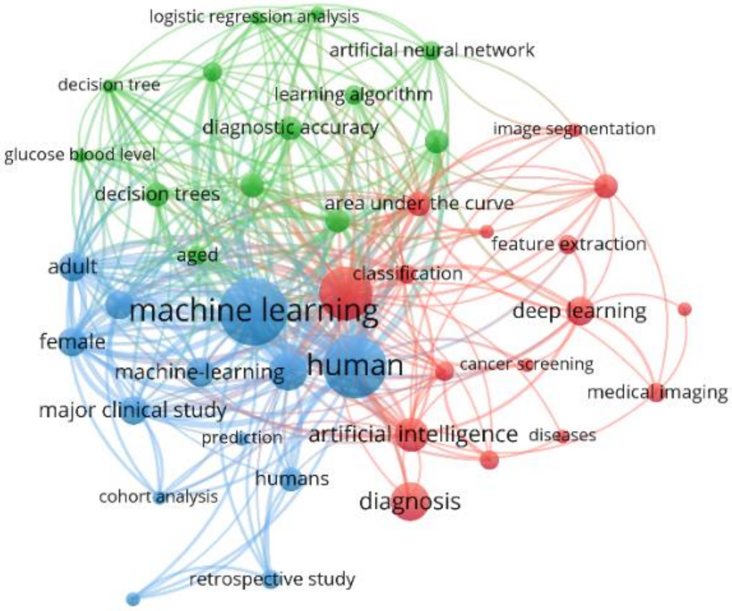

Figure 6: Co-occurrence Keywords Clusters

Figure 6 shows the keyword co-occurrence from the 26 articles selected for this SLR. The VOSviewer software was used to distribute the words into different groups, where each cluster groups keywords based on how frequently they appear together in the same articles. Out of the 26 analyzed papers, 42 key terms were found with a minimum of 3 occurrences across the articles. Notably, the words "machine learning" and "human" stood out with 14 and 13 occurrences, respectively.

B. ALGORITHMIC MODELS BY DISEASE

Table 3 shows a set of algorithmic models used for the prediction of a specific disease in each of the research articles.

Table 3. Algorithmic Models by Disease

REFERENCE	DISEASE	MODEL
Lee et al. [12]	Stroke	Logistic Regression SVC Random Forest XGBoost SVC2 Ensemble (MAX) Ensemble (MIN) Ensemble (MEAN) StrokeClassifier
Gulamali et al. [13]	Intracranial hypertension	CNN
Parola et al. [14]	Oral cancer	CNN
Dubbioso et al. [15]	Dysarthria	Decision Tree
Zhang et al. [16]	Cardiovascular risk	Random Forest Extra Trees Bagging Decision Tree XGBoost LGBM Gradient Boosting Support Vector Machine (SVM)AdaBoost Neural Networks (MLP)
Hong et al. [17]	Kidney disease	SVM Logistic Regression
		Decision Tree KNN Random Forest Gradient Boost AdaBoost XGBoost LightGBM
Zhu et al. [18]	Cancer	Logistic Regression Softmax Ridge SVM KNN Naive Bayes Random Forest AdaBoost CatBoost Extra Trees Ligh Gradient Boosting Gradient Boosting XGBoost
Wu et al. [19]	Gastrointestinal	ConvNext T2T-ViT VGG-19 Conformer HiFuse ShuffleNet V2 GLA-TD
Nillmani et al. [20]	Tuberculosis (TB)	CNN
Moni et al. [21]	Diabetes	Logistic Regression KNN SVM Random Forest Decision Trees Neural Networks (MLP) XGBoost LightGBM
Abdelhafez et al. [22]	Diabetes	Naive Bayes Decision Tree Logistic Regression Random Forest Neural Network SVM
Qasem et al. [23]	Cancer	Gradient Boosting Neural network XGBoost Random Forest Bagging Histgradient boosting

		SGD classifier Logistic egression
Zhang et al. [24]	Acute kidney injury	Gradient Boosting Machines Extremely Randomized Forest Generalized Linear Models Random Forest Deep Learning
Panigrahi et al. [25]	Diabetes mellitus	BMNABC + C4.5 BMNABC + KNN BMNABC + NB BMNABC + RF MNABC + RS BMNABC + ODF(RFE) BMNABC + SVM
Cheng et al. [26]	Axial myopia	XGBoost SVR Linear Regression Stacking Regressor
Fırat et al. [27]	Gastrointestinal	Vanilla ViT Swin Transforme ConvMixer MLPMixer ResNet50 SqueezeNet
Gudigar et al. [28]	Hypertension	ResNet50 ResNet50v2 InceptionV3 DenseNet201 XeeptionNet
Alaraimi et al. [29]	Brain tumor	GLCM+SVM GLCM+ANFIS GLCM+KNN GLCM+RF GLCM+Adaboost
Meena et al. [30]	Chronic kidney disease	KNN RFC DTC GB XGBOOST
Karuppasamy et al. [31]	Breast cancer	CLR CSVM-H CSVM VggNet-16 ResNet-50
Mayrose et al. [32]	Dengue	ML-MLP ML-RF ML-SVM ML-FCM ML-SVM
Bhagubai et al. [33]	Seizure	ChronoNet DeepCNN EEGnet Pathology Dynamics UCLA CDx Neural Engineering Lab Brainify.ai Seizure Hunters
Cabrera Gaytán et al. [34]	Dengue	ANN ANN-ROC DIRECT+ANN-ROC
Li et al. [35]	Cerebral ischemia	U-Net U-Net++ ResU-Net Swin-Unet DCSAU-Net GA-UNet HmsU-Net MCNMF-Unet LeaNet MDU-Net UTAC-Net(Ours)
Chang et al. [36]	Skin cancer	AlexNet+VGGNet GoogleNet+AlexNet ResNet50+InceptionV3 InceptionV3+Xception
		Customized EfficientNet-b4 with ImageNet R-RGB-1 SVM Proposed Hybrid CNN-DenseNet
Hu et al. [37]	Prostate cancer	TPAS

As shown in *Figure 7*, the most recurring prediction algorithms across the set of articles are machine learning models related to decision trees such as RANDOM FOREST and XGBOOST, followed by SVM and the CNN convolutional neural network model.

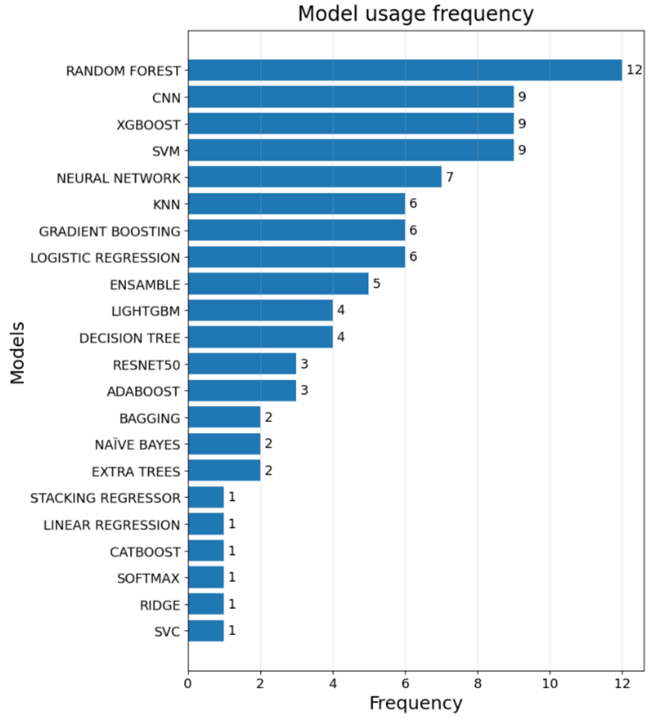

Figure 7: Frequency of use of models

C. *EFFICACY OF MODELS ACCORDING TO DISEASE*

From the total number of study items, 21 different diseases or medical conditions shown in Table 4 were identified, each associated with a specific model such as CNN, RANDOM FOREST, XGBOOST, etc. In each situation, several performance metrics were recorded, such as ACCURACY, ACCURACY, SENSITIVITY, F1-SCORE and AUC.

Table 4. Best performing algorithmic model metrics

DISEASE	MODEL	METRIC
Stroke	StrokeClassifier (Custom)	Accuracy: 0.744 ± 0.009 Precision: 0.743 ± 0.009 F1 score: 0.740 ± 0.010 Kappa: 0.629 ± 0.014
Intracranial hypertension	CNN	Auroc: 0.80 Precision: 0.738 Sensitivity: 0.74 Specificity: 0.73
Oral cancer	CNN	Accuracy: 0.777
Dysarthria	RANDOM FOREST	Wk: 0.8060
Cardiovascular risk	XGBOOST	AUC: 0.902 Accuracy: 0.867 F1-Score: 0.701

Kidney disease	GRADIENT BOOST	Sensitivity: 0.594 Specificity: 0.877 Precision: 0.808 Auroc: 0.826 F1-Score: 0.603
Cancer	CATBOOST	Auc: 0.852 Accuracy: 0.796 Sensitivity: 0.761 Specificity: 0.832
Gastrointestinal	GLA-TD	Accuracy: 0.9063 Precission: 0.8958 Recall: 0.9124 F1-Score: 0.9076
Tuberculosis (TB)	CNN	Accuracy: 0.9929 Precision: 0.9930 Recall: 0.9929 F1-Score: 0.9929 Auc: 0.999
Diabetes	LIGHTGBM	Accuracy: 0.9615 Precission: 0.9677 Recall: 0.9677 F1-Sore: 0.9677 Auroc: 0.9582
Diabetes	RANDOM FOREST	Accuracy: 0.9967 Sensitivity: 0.997 Specificity: 0.983
Cancer	RANDOM FOREST	AUC: 0.717 Accuracy: 0.718 Senstivity: 0.684 Specificity: 0.750
Acute kidney injury	DEEP LEARNING	AUC: 0.830
Diabetes mellitus	RANDOM FOREST	Accuracy = 0.9728 Specificity = 0.9938 Sensitivity = 0.9712
Axial myopia	SVM	MAE = 0.37 ± 0.34 RMSE = 0.82
Gastrointestinal	ResNet50 (CNN)	Accuracy = 0.8744 F1-Score = 0.87 Precision = 0.88 Sensitivity = 0.87
Hypertension	ResNet50 (CNN)	Accuracy = 0.925 Specificity = 0.901 Sensitivity = 0.93 UAC = 0.90
Brain tumor	SVM	Accuracy = 0.9640 Precision = 0.9643 Sensitivity = 0.9770 Specificity = 0.9439 F1-Score = 0.9706
Chronic kidney disease	RANDOM FOREST	Precision = 1.00 Sensitivity = 1.00 F1-Score = 1.00
Breast cancer	ResNet50 (CNN)	Sensitivity = 0.99 Specificity = 0.94 Precision = 0.98 F1-Score = 0.96 AUC = 0.97
Dengue	RANDOM FOREST	Accuracy = 0.9439
Seizure	ChronoNet	AUC = 0.794
Dengue	DIRECT-ANN-ROC	Sensitivity = 0.89 Specificity = 0.99 Ppv = 0.99 Npv = 087 Kapa = 0.87 Ase = 0.02
Cerebral ischemia	UTAC-Net	Dice Coefficient = 0.9175 (IoU %) = 0.8476
Skin cancer	SVM	Accuracy = 0.957
Prostate cancer	DL-TPAS	AUC = 0.82 AUPRC = 0.74

IV. DISCUSSION

Each article used in this systematic literature review has evaluated different types of artificial intelligence algorithms. The most commonly used are machine learning models related to decision trees such as Random Forest [16] [17] [18] [23] [24]; followed by models like XGBoost and SVM. Additionally, the use of Convolutional Neural Networks (CNN) for computerized image analysis in clinical contexts is quite recurrent across various publications [13] [27] [32].

The most recurrent type of prediction problem in the use of AI algorithmic models for disease diagnosis is identified through the metrics that each author has used to measure their models' performance. Although the authors do not explicitly mention it, the metrics used to measure each model's performance, such as AUC, sensitivity, and precision [16], demonstrate that each article is working with classification problems, as they predict a qualitative value. If, alternatively, other types of metrics that are also popular in the machine learning context had been used, such as R^2, MAE, MAPE, or RMSE, the type of problem would be regression [26], where the prediction aims at a continuous numerical value.

In this regard, with respect to accuracy levels, Random Forest shows sensitivity and accuracy results above 90% according to the authors Monit et. al. [21] and Abdelhafez et. al. [22] compared to KNN models, where their performance ranges between 50% and 80% according to Hong et. al. [17] and Meena et. al. [30]. It was also found that CNN can reach performance levels of 99% according to Nillmani [20] in its investigation of computerized image analysis for the diagnosis of tuberculosis.

Additionally, ML and DL models for disease diagnosis have been applied across a diverse range of healthcare institutions, including prestigious academic medical centers in the United States. [13], European university hospitals [14] [15], multiple tertiary hospitals in Asia [17] [18], specialized disease-specific centers [12] [22], regional hospitals and provincial hospitals. Similarly, it is noteworthy that some research has been based on data from national health systems and ministries of health [25], which suggests implementation at a governmental level.

A notable finding among the 26 selected papers is that some authors employ multiple models in combination to make predictions, achieving better results than when using individual algorithms. Lee et al. [12] integrated nine different models to classify stroke etiology in their study, achieving a precision level of 0.74. Meanwhile, Parola et al. [14] combined YOLO, DETR, and R-CNN models into a single ensemble model that reached an accuracy of 0.85 for oral cancer detection. Similarly, Zhang et al. [24] merged four models using AutoML in their work on acute kidney injury diagnosis, obtaining an AUC result of 0.96. These cases demonstrate how ensemble models are transitioning from an innovative framework to a more widespread practice in medical contexts.

It is important to note that, although multiple studies have applied machine learning algorithms for disease diagnosis, ethical concerns remain regarding their use in medical practice. Some authors argue that before AI can be widely adopted in clinical settings, key challenges such as data interpretability and model transparency must first be addressed [23] [25]. As noted by Wei et al., more interpretable AI models, such as Logistic regression, tend to yield lower predictive performance compared to complex models like neural networks. However, while DL algorithms may achieve superior diagnostic accuracy, their inherent lack of explainability prevents clinicians from understanding the underlying decision-making process. This limitation compromises the physician's ability to clearly communicate the rationale for a diagnosis to the patient, thereby undermining the fundamental principle of autonomy and the requirement for truly informed consent [38].

V. CONCLUSION

In this research, the most effective AI models for disease diagnosis were identified, highlighting the use of decision tree-related algorithms such as Random Forest with an average accuracy of 0.91, XGBoost with an average accuracy of 0.87, and convolutional neural networks with an average accuracy of 0.88. According to evaluation metrics such as precision, sensitivity, accuracy, and AUC, we found effective models that can reach 90% effectiveness in the analyzed studies. Thus, the studies conducted demonstrate that Random Forest is highly efficient in detecting dermatological, cardiological, and oncological diseases. This contrasts positively with other machine learning methods such as KNN and SVM. However, difficulties have been encountered in interpreting the results and applying these algorithms in clinical environments with technical limitations; this shows the need to improve the integration of these tools in the healthcare system. These difficulties highlight that for future projects, it is recommended to evaluate the effectiveness of mixed methods that combine Random Forest with conventional methods, as well as conduct longitudinal studies that assess the impact of these models on improving patient prognosis.

REFERENCES

[1] A. Kloczkowski *et al.*, "Machine Learning Models for the Identification of Prognostic and Predictive Cancer Biomarkers: A Systematic Review," *Int. J. Mol. Sci*, vol. 2023, p. 7781, 2023, doi: 10.3390/ijms.

[2] J. Zhao, L. Li, J. Li, and L. Zhang, "Application of artificial intelligence in rheumatic disease: a bibliometric analysis," *Clin Exp Med*, vol. 24, no. 1, Dec. 2024, doi: 10.1007/s10238-024-01453-6.

[3] D. E. Rodriguez-Obregon *et al.*, "Semi-supervised COVID-19 volumetric pulmonary lesion estimation on CT images using probabilistic active contour and CNN segmentation," *Biomed Signal Process Control*, vol. 85, Aug. 2023, doi: 10.1016/j.bspc.2023.104905.

[4] E. J. Ticlavilcainche, M. I. Moreno-Lozano, P. Castañeda, S. Wong-Durand, and A. Oñate-Andino, "Mobile Application Based on Convolutional Neural Networks for Pterygium Detection in Anterior Segment Eye Images at Ophthalmological Medical Centers," *International journal of online and biomedical engineering*, vol. 20, no. 8, pp. 115–138, May 2024, doi: 10.3991/ijoe.v20i08.48421.

[5] A. B. Gibson, R. Robertson, P. M. Urie, and D. Della Corte, "Don't Fear the Artificial Intelligence: A Systematic Review of Machine Learning for Prostate Cancer Detection in Pathology," May 01, 2024, *College of American Pathologists*. doi: 10.5858/arpa.2022-0460-RA.

[6] S. Panja, S. Rahem, C. J. Chu, and A. Mitrofanova, "Big Data to Knowledge: Application of Machine Learning to Predictive Modeling of Therapeutic Response in Cancer," *Curr Genomics*, vol. 22, no. 4, pp. 244–266, Dec. 2020, doi: 10.2174/1389202921999201224110101.

[7] S. Albaradei *et al.*, "Machine learning and deep learning methods that use omics data for metastasis prediction," Jan. 01, 2021, *Elsevier B.V.* doi: 10.1016/j.csbj.2021.09.001.

[8] A. Carriero, L. Groenhoff, E. Vologina, P. Basile, and M. Albera, "Deep Learning in Breast Cancer Imaging: State of the Art and Recent Advancements in Early 2024," Apr. 01, 2024, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/diagnostics14080848.

[9] J. Olveres *et al.*, "What is new in computer vision and artificial intelligence in medical image analysis applications," Aug. 01, 2021, *AME Publishing Company*. doi: 10.21037/qims-20-1151.

[10] L. Muflikhah, A. G. Nurfansepta, F. A. Bachtiar, and D. E. Ratnawati, "High Performance for Predicting Diabetic Nephropathy Using Stacking Regression of Ensemble Learning Method," *International journal of online and biomedical engineering*, vol. 20, no. 8, pp. 149–164, May 2024, doi: 10.3991/ijoe.v20i08.48387.

[11] S. Baral, S. Satpathy, D. P. Pati, P. Mishra, and L. Pattnaik, "A Literature Review for Detection and Projection of Cardiovascular Disease Using Machine Learning," 2024, *European Alliance for Innovation*. doi: 10.4108/eetiot.5326.

[12] H. J. Lee *et al.*, "StrokeClassifier: ischemic stroke etiology classification by ensemble consensus modeling using electronic health records," *NPJ Digit Med*, vol. 7, no. 1, Dec. 2024, doi: 10.1038/s41746-024-01120-w.

[13] F. Gulamali *et al.*, "Derivation, external and clinical validation of a deep learning approach for detecting intracranial hypertension," *NPJ Digit Med*, vol. 7, no. 1, Dec. 2024, doi: 10.1038/s41746-024-01227-0.

[14] M. Parola, F. A. Galatolo, G. La Mantia, M. G. C. A. Cimino, G. Campisi, and O. Di Fede, "Towards explainable oral cancer recognition: Screening on imperfect images via Informed Deep Learning and Case-Based Reasoning," *Computerized Medical Imaging and Graphics*, vol. 117, Oct. 2024, doi: 10.1016/j.compmedimag.2024.102433.

[15] R. Dubbioso *et al.*, "Precision medicine in ALS: Identification of new acoustic markers for dysarthria severity assessment," *Biomed Signal Process Control*, vol. 89, Mar. 2024, doi: 10.1016/j.bspc.2023.105706.

[16] K. Zhang *et al.*, "Machine-learning-based models to predict cardiovascular risk using oculomics and clinic variables in KNHANES," *BioData Min*, vol. 17, no. 1, Dec. 2024, doi: 10.1186/s13040-024-00363-3.

[17] S. H. Hong *et al.*, "Machine learning models for predicting the onset of chronic kidney disease after surgery in patients with renal cell carcinoma," *BMC Med Inform Decis Mak*, vol. 24, no. 1, Dec. 2024, doi: 10.1186/s12911-024-02473-8.

[18] H. Zhu *et al.*, "A machine learning prediction model for cancer risk in patients with type 2 diabetes based on clinical tests," *Technology and Health Care*, vol. 32, no. 3, pp. 1431–1443, May 2024, doi: 10.3233/THC-230385.

[19] P. Wu, H. Li, L. Hu, J. Ge, and N. Zeng, “A Local-Global Attention Fusion Framework with Tensor Decomposition for Medical Diagnosis,” *IEEE/CAA Journal of Automatica Sinica*, vol. 11, no. 6, pp. 1536–1538, Jun. 2024, doi: 10.1109/JAS.2023.124167.

[20] Nillmani, V. Sharma, S. K. Gupta, and K. K. Shukla, “Deep learning models for tuberculosis detection and infected region visualization in chest X-ray images,” *Intelligent Medicine*, vol. 4, no. 2, pp. 104–113, May 2024, doi: 10.1016/j.imed.2023.06.001.

[21] M. A. Moni, N. Nipa, M. H. Riyad, S. Satu, Walliullah, and K. C. Howlader, “Clinically adaptable machine learning model to identify early appreciable features of diabetes,” *Intelligent Medicine*, vol. 4, no. 1, pp. 22–32, Feb. 2024, doi: 10.1016/j.imed.2023.01.003.

[22] H. A. Abdelhafez and A. A. Amer, “Machine Learning Techniques for Diabetes Prediction: A Comparative Analysis,” *Journal of Applied Data Sciences*, vol. 5, no. 2, pp. 792–807, May 2024, doi: 10.47738/jads.v5i2.219.

[23] J. Di *et al.*, “Utility of artificial intelligence in a binary classification of soft tissue tumors,” *J Pathol Inform*, vol. 15, Dec. 2024, doi: 10.1016/j.jpi.2024.100368.

[24] R. Zhang, M. Yin, A. Jiang, S. Zhang, X. Xu, and L. Liu, “Automated machine learning for early prediction of acute kidney injury in acute pancreatitis,” *BMC Med Inform Decis Mak*, vol. 24, no. 1, Dec. 2024, doi: 10.1186/s12911-024-02414-5.

[25] R. Panigrahi *et al.*, “Optimized Forest Framework with A Binary Multineighborhood Artificial Bee Colony for Enhanced Diabetes Mellitus Detection,” *International Journal of Computational Intelligence Systems*, vol. 17, no. 1, Dec. 2024, doi: 10.1007/s44196-024-00598-2.

[26] H. Cheng *et al.*, “Systematic evaluation of machine learning-enhanced trifocal IOL power selection for axial myopia cataract patients,” *Comput Biol Med*, vol. 173, May 2024, doi: 10.1016/j.compbiomed.2024.108245.

[27] H. Fırat, A. A. Demirbaş, and H. Üzen, “Spatial-attention ConvMixer architecture for classification and detection of gastrointestinal diseases using the Kvasir dataset,” *Health Inf Sci Syst*, vol. 12, no. 1, Dec. 2024, doi: 10.1007/s13755-024-00290-x.

[28] N. A. K. U. R. J. S. M. A. I. M. A. P. U. R. A. Anjan Gudigar, “Directional-Guided Motion Sensitive Descriptor for Automated Detection of Hypertension Using Ultrasound Images,” *IEEE_Access*, 2024.

[29] S. Alaraimi, I. Al Naimi, S. Manic, N. Al Hinai, and S. Al Shukaili, “Enhancing Brain Tumor Assessment: A Comprehensive Approach using Computerized Diagnostic Tool and Advanced MRI Techniques,” in *Procedia Computer Science*, Elsevier B.V., 2024, pp. 3350–3368. doi: 10.1016/j.procs.2024.04.316.

[30] K. Meena, K. Hema, and R. Pandian, “Analyze the impact of feature selection techniques in the early prediction of CKD,” *International Journal of Cognitive Computing in Engineering*, vol. 5, pp. 66–77, Jan. 2024, doi: 10.1016/j.ijcce.2023.12.002.

[31] A. D. Karuppasamy, A. Abdesselam, R. Hedjam, H. zidoum, and M. Al-Bahri, “Feed-forward networks using logistic regression and support vector machine for whole-slide breast cancer histopathology image classification,” *Intell Based Med*, vol. 9, Jan. 2024, doi: 10.1016/j.ibmed.2023.100126.

[32] H. Mayrose, N. Sampathila, G. Muralidhar Bairy, T. Nayak, S. Belurkar, and K. Saravu, “An Explainable Artificial Intelligence Integrated System for Automatic Detection of Dengue From Images of Blood Smears Using Transfer Learning,” *IEEE Access*, vol. 12, pp. 41750–41762, 2024, doi: 10.1109/ACCESS.2024.3378516.

[33] M. Bhagubai, L. Swinnen, E. Cleeren, W. Van Paesschen, M. De Vos, and C. Chatzichristos, “Towards Automated Seizure Detection with Wearable EEG - Grand Challenge,” *IEEE Open Journal of Signal Processing*, vol. 5, pp. 717–724, 2024, doi: 10.1109/OJSP.2024.3378604.

[34] D. A. Cabrera Gaytán *et al.*, “Effectiveness of a diagnostic algorithm for dengue based on an artificial neural network,” *Digit Health*, vol. 10, Jan. 2024, doi: 10.1177/20552076241237691.

[35] W. Li and W. Zhang, “UTAC-Net: A Semantic Segmentation Model for Computer-Aided Diagnosis for Ischemic Region Based on Nuclear Medicine Cerebral Perfusion Imaging,” *Electronics (Switzerland)*, vol. 13, no. 8, Apr. 2024, doi: 10.3390/electronics13081466.

[36] H. T. Chang, A. De, and N. Mishra, “An approach to the dermatological classification of histopathological skin images using a hybridized CNN-DenseNet model,” *PeerJ Comput Sci*, vol. 10, 2024, doi: 10.7717/peerj-cs.1884.

[37] L. Hu *et al.*, “Development and Validation of a Deep Learning Model to Reduce the Interference of Rectal Artifacts in MRI-based Prostate Cancer Diagnosis,” *Radiol Artif Intell*, vol. 6, no. 2, Mar. 2024, doi: 10.1148/ryai.230362.

[38] S. Wei *et al.*, “Application of Machine Learning for Patients With Cardiac Arrest: Systematic Review and Meta-Analysis,” *J Med Internet Res*, vol. 27, 2025, doi: 10.2196/67871.