Indonesian Journal of Electrical Engineering and Computer Science
Vol. 35, No. 1, July 2024, pp. 418~427
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v35.i1.pp418-427

Reviewing chronic ailments: predicting diseases with a multi-symptom approach

Aicha Oussous, Abderrahmane Ez-Zahout, Soumia Ziti
Department of Computer Science, Faculty of Sciences, Mohammed V University in Rabat, Rabat, Morocco

Article Info

Article history:

Received Dec 6, 2023
Revised Feb 15, 2024
Accepted Mar 20, 2024

Keywords:

Chronic diseases
Ensemble model
Hybrid model
Machine learning
Symptoms

ABSTRACT

The integration of machine learning (ML) techniques is now indispensable in healthcare, especially in addressing the challenges posed by chronic illnesses, which present a significant global health concern due to their unpredictable nature. This study compares ML techniques employed in the diagnosis and treatment of chronic conditions such as diabetes, liver disease, thyroid disease, breast cancer, heart disease, Alzheimer's disease, and others. Two primary criteria guided the selection of diseases under investigation. Firstly, those extensively studied with ML methods, and secondly, those leveraging ML models to resolve issues or yield promising results. The research concludes that in real-time clinical practice, there is no universally proven method for selecting the optimal course of action due to each method's unique advantages and disadvantages. While a hybrid technique may exhibit slightly slower speed growth, it holds the potential to enhance the accuracy and performance of a model.

Corresponding Author:

Aicha Oussous
Department of Computer Science, Faculty of Sciences, Mohammed V University in Rabat
Rabat, Morocco
Email: aicha_oussous@um5.ac.ma

1. INTRODUCTION

Machine learning (ML) is an artificial intelligence (AI) method in which a machine learns and improves it is performance based on previous experiences. Healthcare is currently being influenced by ML algorithms. Health data is highly sensitive, and any error might jeopardize a person's life. Humans are unable to process data quickly using traditional methods. In these circumstances, ML techniques are employed to determine illness patterns and causes. The integration of new technologies, such as ML, into healthcare facilitates its development [1]. ML techniques are also employed in a variety of applications, including illness diagnosis, drug detection, and assistive technology. Accuracy, decision-making, rapid and powerful processing, managing complicated data, and cost-effectiveness are all advantages of ML algorithms. Several tasks have shown promise for the use of ML approaches, including the classification of interstitial lung diseases, such as the segmentation of brain tumours, the identification of body parts in medical pictures, and the detection of lung nodules [2]. ML models have already surpassed human performance in disciplines such as clinical dermatology, ophthalmology, radiology, and pathology [2]. Additionally, it will be possible to predict patient outcomes, identify chronic diseases, and reduce death rates brought on by these diseases using ML models.

This study focuses on the prediction of chronic illnesses, one of the primary contributors to decreased quality of life and increased healthcare costs. Through frequent hospitalizations, disabilities, and treatment costs, chronic illnesses impose significant burdens on individuals and healthcare systems. According to [3], the cost of therapies for these diseases equals more than 70% of a patient's income.

The direct costs of chronic illnesses to healthcare systems in the United States are close to US$214 billion annually. Furthermore, lost productivity at work due to chronic illnesses costs US$138 billion. According to Delpino *et al.* [4], the costs associated with chronic illnesses are much greater in low- and middle-income nations than in high-income ones, as indicated by [5]'s findings.

AI's potential for use in a variety of industries, including healthcare, has improved in recent years. According to [5], information technology platforms are already in place, along regional medical and public health collaboration, as well as individual electronic health records, to develop the fundamental components for AI-based services for chronic illness management systems. This study examines ML methods employed for diagnosing and treating chronic diseases such as heart disease, Alzheimer's, diabetes, liver disease, thyroid disease, breast cancer, and more. Refer to Table 1 in the appendix for a concise summary of the incorporated studies. The organization of the paper is as follows: section 2 outlines the methods used; section 3 delves into the results; section 4 addresses challenges and potential future work; and the paper wraps up with a concise summary in section 5.

2. METHODS

The goal of this study is to review the prediction of chronic diseases. For that, we have collected various papers from different sources (Science Direct, Google Scholar, Springer Nature, Springer databases, and IEEE Xplore) by using expressions like "Chronic diseases using ML", "Novelty ML and deep learning in disease prediction", or "Chronic diseases categorization using machine learning". There were 443 documents found after the search, as shown in Figure 1. The amount 265 papers were deleted after studying the abstracts. Doctorate dissertations, reports, theses issued in languages other than English, and studies that do not predict the incidence of chronic diseases are excluded from our review, 112 objects were deleted in this examination. After reviewing the complete text, 35 papers were deleted because they used similar methodologies or had already been presented elsewhere, and 66 publications were selected for detailed review. In the end, this study looked at 31 articles.

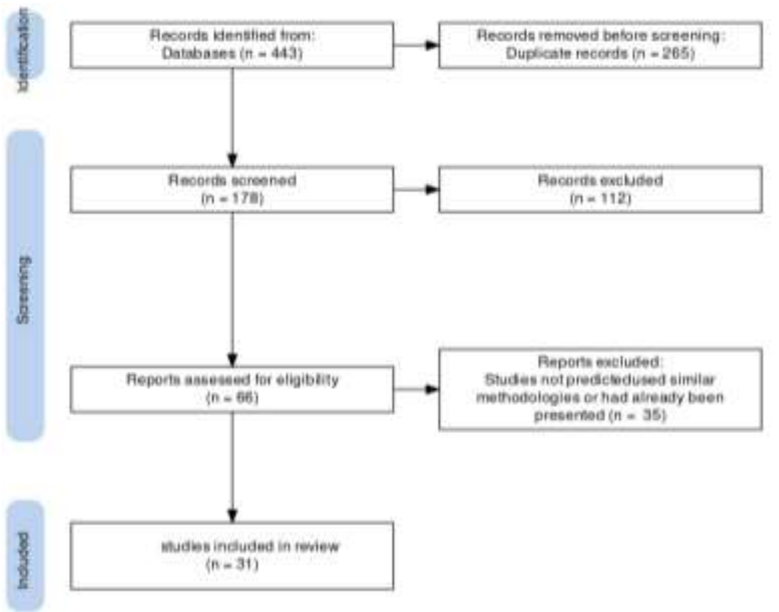

Figure 1. PRISMA diagram showing the included studies chosen for the review

Table 1 in the APPENDIX provides a summary of the evaluated studies in this publication, outlining details such as illness type, dataset, employed algorithms, and metrics used for evaluation (including accuracy, precision, sensitivity, specificity, area under the curve (AUC), and F1-score) for each study. The reviewed research incorporates various methods like decision tree (DT), Naïve Bayes (NB), k-nearest neighbor (KNN), logistic regression (LR), support vector machine (SVM), and random forest (RF). These algorithms are not only applied to standard datasets across multiple illnesses but are also anticipated to play an increasingly significant role in medical practice in the near future. These algorithms are also valuable for categorizing and diagnosing chronic disorders.

Based on our current understanding, this research serves as the initial evaluation, examining a broad array of metrics such as accuracy, precision, sensitivity, specificity, AUC, and F1-score relevant to ML algorithms predicting chronic illnesses. Our primary findings confirm that ML algorithms demonstrate a significant ability to predict chronic illnesses with high accuracy. The reviewed articles covered a range of chronic diseases, including heart disease, chronic kidney disease, diabetes, Alzheimer’s disease, thyroid disease, liver disease, breast cancer, cerebral infection, and hypertension.

3. DISCUSSION AND RESULTS

Diverse algorithms, encompassing KNN, LR, DT, SVM, NB, and RF, were applied to varied datasets and features in [6]. The study revealed that the DT algorithm outperformed the SVM method, contrasting with [7], which observed the reverse, with the SVM method surpassing the DT algorithm. In the context of liver disease identification, [8] explored six alternative methods, with LR exhibiting the highest accuracy among them. Mishra *et al.* [9], LR achieved an accuracy of 98.95%, RF reached 99.75%, and the hybridization of LR and RF attained the peak accuracy at 99.83%. Post-feature selection in [10], the KNN algorithm outperformed all other methods; the authors emphasize that feature selection is a pivotal step in every ML model. Conversely, in [11], the SVM technique coupled with the recursive feature elimination (RFE) feature selection technique yielded the highest accuracy. The assertion is that the advantages of feature selection include a reduction in overfitting, improved accuracy, and shorter training time. Bharti *et al.* [12], the application of the DL approach to the original dataset yielded an accuracy of 76.7%. However, through feature selection and outlier detection, the accuracy significantly improved to 94.2%. Asnaoui [13], combining ResNet50, MobileNet V2, and InceptionResNet V2 achieved an accuracy of 95.09%, surpassing individual accuracies of InceptionResNet V2 (94.50%), MobileNet V2 (93.73%), and ResNet50 (93.73%). This underscores the effectiveness of combining DL algorithms in enhancing overall model accuracy. Reddy *et al.* [14], the accuracy of individual methods such as RF, DT, Adaboost classifier, KNN, and LR was found to be lower compared to the combined accuracy of all these algorithms, resulting in an 80% accuracy. The study emphasizes that hybridization and the number of approaches in ensemble ML algorithms significantly impact the model’s accuracy. The researchers [15]–[18] explored hybrid deep learning approaches, combining the CNN algorithm with another algorithm. The outcomes revealed varying accuracy levels depending on the specific procedures employed. The careful selection of algorithms for merging is crucial, as it directly influences the overall performance of the model. The study concludes that all methods demonstrate effective performance even with small datasets.

Based on this study, we can derive the following conclusions and identify both favorable and unfavorable outcomes. First, the choice of algorithm(s) for disease prediction is contingent upon the specifics of the task and the available data. Optimal strategy selection may require testing various algorithms to assess their effectiveness. Second, predicting the best-performing algorithm without testing is challenging, given that different algorithms may exhibit distinct behaviors on varied datasets and attributes. It is advisable to test a range of algorithms, evaluating performance measures such as accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC), to identify the most suitable algorithm(s) for the given problem. Third, the performance of a predictive model is influenced by multiple factors, including the chosen algorithm(s), data quality and volume, feature engineering and preprocessing methods, selected hyperparameters, and the evaluation methodology. Thoroughly assessing model performance and considering all relevant criteria is crucial before deploying the final model(s). Fourth, the accuracy of a ML algorithm is affected by the dataset size, quantity and quality of features, problem difficulty, chosen methods, and selected hyperparameters. Fifth, in comparative studies, incorporating various datasets is advantageous. However, ensuring meaningful and useful outcomes requires careful evaluation of the quality and relevance of the data. Last, while combining algorithms may enhance results in some cases, success in ensemble learning depends on several parameters, including the quality and diversity of individual algorithms, the combination and weighting techniques employed, and the nature of the problem addressed. An ensemble may not yield substantial benefits if individual algorithms are strongly correlated or exhibit similar biases, and the approach used, such as simple averaging or majority voting, may be inappropriate.

This following study identify positive and negative outcomes. Positive outcomes: i) accuracy: when dealing with huge amounts of data, ML models can produce predictions that are more accurate than those produced by conventional statistical models; ii) scalability: ML models are capable of processing enormous volumes of data and may be trained using data from a variety of sources, including electronic health records and medical imaging; iii) automation: predictions may be made more quickly and with less effort by using ML models, which can automate the process; and iv) better patient outcomes can result from earlier intervention and the accurate and early diagnosis of chronic diseases. Negative outcomes: i) data bias: ML algorithms are only as accurate as the data used to train them, and data bias can lead to erroneous predictions;

ii) overfitting: overfitting can happen when a model has too many parameters and is overly complicated. As a result, performance results on training data may be overly optimistic, while results on unobserved data may be subpar; iii) inability to be interpreted: some ML models, such as deep learning models, can be challenging to grasp how the model arrived at a specific prediction; and iv) cost: creating and deploying ML models can be costly and necessitate specific training and expertise.

4. CHALLENGE AND POTENTIAL FUTURE WORK

Based on the preceding section, various chronic diseases pose distinct challenges, encompassing, i) lack of adequate data: finding adequate data to train ML models is one of the main obstacles to disease prediction. It is challenging to produce precise forecasts because the quantity and quality of data vary substantially among various diseases; ii) disease complexity: many illnesses are complicated and multifactorial, meaning that various genetic, environmental, and lifestyle variables contribute to their development. An accurate illness prediction requires a thorough understanding of the underlying biology and a large dataset with all pertinent variables; iii) data bias: the potential for data bias is another difficulty in disease prediction. This can happen when specific variables are overrepresented or underrepresented in the data, or when the data used to train the model does not represent the researched population; iv) interpreting results: even when a ML model produces accurate predictions, the interpretation of the results can be challenging. A deep understanding of the underlying biology and statistical analysis is necessary to comprehend the elements that go into prediction and how these factors interact with one another; v) generalization to new populations: ML models trained on a single population may not generalize well to different groups. This can be especially difficult for diseases with varying risk factors or prevalence rates in various populations; and vi) ethical and legal issues: utilizing ML to forecast diseases presents several ethical and legal concerns. For instance, there may be issues with data privacy, informed consent, and the use of private medical data.

We will guide our future work according to the conclusions drawn from the preceding analysis, which indicated that: i) the amalgamation of data from diverse sources yields more impactful and practical solutions [19]; ii) the application of feature extraction and selection techniques is crucial in ML [20]; and iii) enhanced accuracy in predictions is achieved through the utilization of hybrid ML algorithms [21]–[23]. We will implement these insights by following the steps outlined below, as illustrated in Figure 2:

- Collect the data sets: we will focus on gathering data related to the disorders examined, encompassing diabetes, cancer, thyroid issues, liver conditions, kidney diseases, Alzheimer's, hypertension, and cardiovascular ailments. Additionally, diseases not covered in the initial study may be included.
- Prepare the data: normalize the data, eliminate lower-ranked values, remove duplicate entries, and address missing values.
- Extract and select features: identify common indicators across all diseases, recognize distinctive symptoms for each disease, compile all symptoms, and construct a new dataset encompassing all symptoms and selected features.
- Select ML techniques: drawing from the ML algorithms explored in the study, we will: i) determine the optimal combination strategy, ii) select algorithms that produce the most favorable results; and iii) combine algorithms using the chosen strategies.
- Predict diseases: after assessing the chosen ML algorithms using relevant metrics, the model will predict the likelihood of the existence of a disease and estimate it is severity percentage.

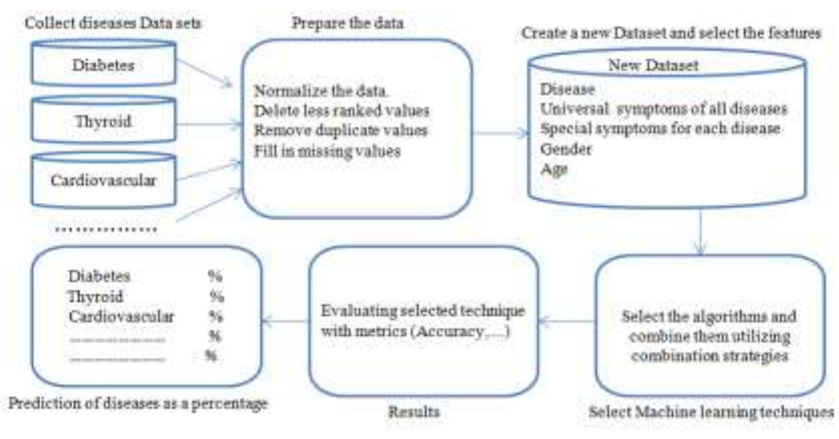

Figure 2. Ensemble model based on the symptoms

5. CONCLUSION

When cutting-edge technologies like ML are integrated into the medical sector, they empower healthcare professionals with tools to analyze disease-related data. Consequently, ML algorithms play a crucial role in facilitating the early detection of diseases. This research extensively examined various ML algorithms for predicting illnesses, utilizing standard datasets across conditions such as liver diseases, heart ailments, breast cancer, and others. Researchers employed ML algorithms to diagnose illnesses, relying on a set of outcomes. A thorough review of preceding articles focusing on illness prediction models revealed that certain algorithms, including NB, SVM, KNN, RF, and DT, exhibited excellent accuracy. However, it was noted that the accuracy of a given method could vary across datasets. This variability is influenced by significant parameters such as the nature of the datasets, the process of feature selection, and the number of features considered, all of which impact the model's accuracy and overall performance. To address these considerations and enhance accuracy and performance, the forthcoming study will introduce a novel approach: constructing a unified ensemble model based on the signs and symptoms associated with each illness. This innovative paradigm aims to further advance the field of disease prediction.

APPENDIX

Table 1. Summary of the included studies

Ref	Diseases/ Year	Dataset	Used Algorithms	Accuracy (%)	Precision (%)	Sensitivity (Recall) (%)	Specificity (%)	AUC (%)	F1-measure (%)
[24]	2023 Heart	Heart disease dataset	LR	85	82	84			84.5
			SVR	84	83	83			84
			RF	98.6	97.8	98.1			98.4
[25]	2023 Chronic Kidney	Kaggle	DT	75					
			LR	80					
			NB	76.6					
[26]	2022 Alzheimer's disease	From the gene expression omnibus (GEO)	K-NN	89					
			NB	85					
			SVM	96					
[27]	2022 Heart	From UCI	ANN	85.8		85.4			
			SVM	89.1		89.1			
[28]	2022 COVID-19	Obtained from the registry of Ayatollah Taleghani Hospital, Abadan city, Iran,	RF	95.03	94.23	90.70	95.10		
			XGBoost	94.25	92.43	90.89	95.01		
			KNN	89.56	80.11	97.38	82.15		
			multi-layer perceptron (MLP)	91.25	87.19	90.81	91.07		
			LR	91.23	83.94	91.45	84.47		
			J48 decision tree	92.17	89.97	87.77	94.47		
			NB	87.47	81.32	90.44	84.31		
[29]	2021 Breast Cancer	The curated breast imaging subset of DDSM (CBISDDSM)	Extreme Gradient Boosting (XGBoost)		64.11	64.09		68.29	64.10
			VGG-16		64.05	64.06		68.22	64.05
[30]	2021 Diabetes		LR	84					91.11
			SVM	84					91.3
			RF	79.6					88.75
[31]	2021 Alzheimer		LR	88.24					
			NB	74.65					
			DL	78.32					
			K-NN	43.26					
			DT	74.22					
[32]	2020 Breast Cancer	Wisconsin Breast cancer patient's dataset	K-NN	83.33	96.58				
[33]	2020 Diabetes	From the ML repository	{With Rough K Means}						
			NB	80.55	90	80.14	80.14		84.78
			SVM	77.78	88.19	77.24	78.87		82.35
			RF	77.20	56.9	56.9	69.05		62.06
			K-NN	71.30	77.08	79.29	70.67		78.17

Table 1. Summary of the included studies *(continued...)*

Ref	Diseases/ Year	Dataset	Used Algorithms	Accuracy (%)	Precision (%)	Sensitivity (Recall) (%)	Specificity (%)	AUC (%)	F1-measure (%)
[33]	Breast cancer	Kaggle	NB	94.44	95.92	94.95	93.65		95.43
			SVM	97.53	97.27	99.07	94.44		98.17
			RF	96.30	95	93.09	96.01		93.09
			K-NN	85.80	84.21	95.05	70.49		89.30
	Kidney	Kaggle	NB	98.11	96.15	96.43	6.15		98.04
			SVM	100	100	100	100		100
			RF	100	100	100	100		100
			K-NN	84.91	92.59	80.65	90.91		86.21
[34]	2020 Diabetes	UCI ML repository	RF	68	70	69			69
			DT	55	58	60			59
			Adaboost classifier	67	68	70			69
			K-NN	65	67	65			66
			LR	77	82	74			78
[35]	2020 Heart	University of California (UCI) heart disease Cleveland	Framework consisting of factor Analysis of Mixed Data (FAMD)+ RF	93.44		89.28	96.96	93.12	92.59
			FAMD+ LR	91.80		92.85	90.90	91.88	91.22
			FAMD+KNN	9.16		92.85	87.87	90.36	89.65
			FAMD+DT	81.96		71.42	90.90	81.16	78.43
			FAMD+SVM	91.80		100	84.84	92.42	91.80
[36]	2020 Heart	Two datasets (Statlog and Cleveland)	DBSCAN	98.40					
			SMOTEENN						
			XG BOOST	95.90					
[37]	2020 Breast Cancer	From UCI website	SVM	98					
			Artificial Neural Network	98					
[38]	2020 Alzheimer	Alzheimer's disease Neuroimaging Initiative (ADNI)	SVM	85					
			DT	83					
[39]	2020 Cardio vascular		SVM			61	44	54	
			RF			68	63	68	
			Neural network			76	57	75.3	
			LR			74	57	74.8	
			K-NN			76	60	75.2	
			Gradient boosting machine			74	59	73.7	
	Chronic kidney		SVM			86	65	84.8	
			RF			81	80	89.5	
			Neural network			84	80	90.1	
			LR			87	78	90.5	
			K-NN			81	77	86.6	
			Gradient boosting machine			86	80	90.3	
	Diabetes		SVM			64	48	60.6	
			RF			72	64	73.9	
			Neural network			78	62	76.4	
			LR			74	63	76.8	
			K-NN			82	58	75.8	
			Gradient boosting machine			67	68	76	
	Hypertension		SVM			85	60	78	
			RF			80	63	76.5	
			Neural network			83	58	77.5	
			LR			80	60	77	
			K-NN			61	81	76.8	
			Gradient boosting machine			84	56	76.7	

Table 1. Summary of the included studies *(continued...)*

Ref	Diseases/ Year	Dataset	Used Algorithms	Accuracy (%)	Precision (%)	Sensitivity (Recall) (%)	Specificity (%)	AUC (%)	F1-measure (%)
[40]	2019 Alzheimer	Alzheimer's Disease Neuro imaging Initiative (ADNI)	LR	98.12		90	95		
			DT	97.02		83	84		
			SVM	97		91	90		
[41]	2019 Heart	Kaggle	Random Over sampling: SVM						
			Synthetic Minority Oversampling: RF	99	99.7	100			
			Adaptive synthetic	91.3	93	89			
			Sampling approach: RF	90.3	93	87			
[42]	2019 Heart	From UCI ML repository	NB	87					
			DT	91					
[43]	2019 Diabetes		(Class=0)						
			DT	85	78	71			74
			SVM	77.3	70	100			82
			NB	77	82	79			80
			Artificial Neural Network	82	70	100			82
[44]	2019 Liver	From the UCI ML Repository	LR	75	91	78	47		84
			RF	74	85	81	50		83
			DT	69	77	79	48		78
			SVM	64	69	88	21		77
			K-NN	62	22	76	35		74
			NB	53	36	100	46		53
[45]	2019 Thyroid	From UCI ML repository	DT	99.46	99	99			99
			RF	99.30	99	99			99
			SVM	96.25	96	96			96
			Multilayer Feed forward	95.17	91	95			91
			LR	97.50	97	97			97
[46]	2019 Breast Cancer	The Wisconsin Breast Cancer (Original)	k=10 cross validation						
			SVM	96. 99	97	97			
			* SMO: *LibSVM	95.70	96	95.7			
			Artificial Neural Network *MLP	95.44	95.4	95.4			
			*Voted Perceptron	90.98	91.9	91			
[47]	2018 Heart	UCI repository, Kaggle in the dataset.	DT	64	65	64			65
			RF	65	65	66			65
			SVM	69	60	69			61
[47]	Diabetes		DT	76	76	76			75
			RF	98	70	71			71
			SVM	82.46	82	82			82
	Liver		DT	69.81	70	70			70
			RF	83	84	83			83
			SVM	75.47	81	75			74
[48]	2017 cerebral infarction	From real-life hospitals in central China in 2013–2015	CNN-UDRP	94.2		98.08			
			CNN-MDRP	94.8	99.9				

Table 1. Summary of the included studies *(continued...)*

Ref	Diseases/ Year	Dataset	Used Algorithms	Accuracy (%)	Precision (%)	Sensitivity (Recall) (%)	Specificity (%)	AUC (%)	F1-measure (%)
[49]	2017 Heart	Routine clinical data of 378,256 patients from UK family practices	RF			65.3		74.5	
			LR			67.1		76	
			Gradient boosting machines			67.5		76.1	
			Neural networks			67.5		76.4	
[50]	2016 Diabetes	From the Center for ML and Intelligent Systems at UCI.	NB	65.07					
			SVM	87.32					
			DT	87.46					
			Artificial Neural Networks	76.2					
	Heart	From UCI (University of California, Irvine C.A).	NB	93.85					
			SVM	95.2					
			DT	92.59					
			Artificial Neural Networks	94.27					

REFERENCES

[1] A. Oussous, A. Ez-Zahout, S. Ziti, and A. Oussous, "An overview of the most efficient methods for predicting healthcare disorders," in *AIP Conference Proceedings*, 2023, vol. 2814, no. 1, doi: 10.1063/5.0150025.

[2] N. G. Maity and S. Das, "Machine learning for improved diagnosis and prognosis in healthcare," Mar. 2017, doi: 10.1109/AERO.2017.7943950.

[3] R. Alanazi, "Identification and prediction of chronic diseases using machine learning approach," *Journal of Healthcare Engineering*, pp. 1–9, Feb. 2022, doi: 10.1155/2022/2826127.

[4] F. M. Delpino, K. Costa, S. R. Farias, A. D. P. C. Filho, R. A. Arcêncio, and B. P. Nunes, "Machine learning for predicting chronic diseases: a systematic review," *Public Health*, vol. 205, pp. 14–25, Apr. 2022, doi: 10.1016/j.puhe.2022.01.007.

[5] G. Yu *et al.*, "Improving chronic disease management for children with knowledge graphs and artificial intelligence," *Expert Systems with Applications*, vol. 201, Sep. 2022, doi: 10.1016/j.eswa.2022.117026.

[6] S. Raisinghani, R. Shamdasani, M. Motwani, A. Bahreja, and P. R. N. Lalitha, "Thyroid prediction using machine learning techniques," in *Communications in Computer and Information Science*, vol. 1045, Springer Singapore, 2019, pp. 140–150.

[7] H. A. U. Rehman, C. Y. Lin, Z. Mushtaq, and S. F. Su, "Performance analysis of machine learning algorithms for thyroid disease," *Arabian Journal for Science and Engineering*, vol. 46, no. 10, pp. 9437–9449, 2021, doi: 10.1007/s13369-020-05206-x.

[8] A. K. M. S. Rahman, F. M. J. M. Shamrat, Z. Tasnim, J. Roy, and S. A. Hossain, "A comparative study on liver disease prediction using supervised machine learning algorithms," *International Journal of Scientific and Technology Research*, vol. 8, no. 11, pp. 419–422, 2019.

[9] S. Mishra, P. K. Mallick, H. K. Tripathy, A. K. Bhoi, and A. González-Briones, "Performance evaluation of a proposed machine learning model for chronic disease datasets using an integrated attribute evaluator and an improved decision tree classifier," *Applied Sciences (Switzerland)*, vol. 10, no. 22, pp. 1–35, Nov. 2020, doi: 10.3390/app10228137.

[10] P. Duggal and S. Shukla, "Prediction of thyroid disorders using advanced machine learning techniques," in *Proceedings of the Confluence 2020-10th International Conference on Cloud Computing, Data Science and Engineering*, Jan. 2020, pp. 670–675, doi: 10.1109/Confluence47617.2020.9058102.

[11] D. Jamkhandikar and N. Priya, "Thyroid disease prediction using feature selection and machine learning classifiers," *The International journal of analytical and experimental modal analysis*, vol. 12, pp. 175–180, 2020.

[12] R. Bharti, A. Khamparia, M. Shabaz, G. Dhiman, S. Pande, and P. Singh, "Prediction of heart disease using a combination of machine learning and deep learning," *Computational Intelligence and Neuroscience*, pp. 1–11, Jul. 2021, doi: 10.1155/2021/8387680.

[13] K. El Asnaoui, "Design ensemble deep learning model for pneumonia disease classification," *International Journal of Multimedia Information Retrieval*, vol. 10, no. 1, pp. 55–68, Feb. 2021, doi: 10.1007/s13735-021-00204-7.

[14] G. T. Reddy *et al.*, "An Ensemble based machine learning model for diabetic retinopathy classification," Feb. 2020, doi: 10.1109/ic-ETITE47903.2020.235.

[15] A. Baccouche, B. Garcia-Zapirain, C. C. Olea, and A. Elmaghraby, "Ensemble deep learning models for heart disease classification: A case study from Mexico," *Information (Switzerland)*, vol. 11, no. 4, Apr. 2020, doi: 10.3390/INFO11040207.

[16] S. L. Oh, E. Y. K. Ng, R. S. Tan, and U. R. Acharya, "Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats," *Computers in Biology and Medicine*, vol. 102, pp. 278–287, Nov. 2018, doi: 10.1016/j.compbiomed.2018.06.002.

[17] D. K. Atal and M. Singh, "Arrhythmia classification with ECG signals based on the optimization-enabled deep convolutional neural network," *Computer Methods and Programs in Biomedicine*, vol. 196, Nov. 2020, doi: 10.1016/j.cmpb.2020.105607.

[18] Z. Zheng, Z. Chen, F. Hu, J. Zhu, Q. Tang, and Y. Liang, "An automatic diagnosis of arrhythmias using a combination of CNN and LSTM technology," *Electronics (Switzerland)*, vol. 9, no. 1, Jan. 2020, doi: 10.3390/electronics9010121.

[19] C. Chola, P. Mallikarjuna, A. Y. Muaad, J. V. B. Benifa, J. Hanumanthappa, and M. A. Al-antari, "A hybrid deep learning approach for COVID-19 diagnosis via CT and X-ray medical images," in *IOCA 2021*, Sep. 2022, p. 13, doi: 10.3390/ioca2021-10909.

[20] K. Deepika and S. Seema, "Predictive analytics to prevent and control chronic diseases," in *Proceedings of the 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology, iCATccT 2016*, 2017, pp. 381–386, doi: 10.1109/ICATCCT.2016.7912028.

[21] M. A. Islam, M. Z. H. Majumder, and M. A. Hussein, "Chronic kidney disease prediction based on machine learning algorithms," *Journal of Pathology Informatics*, vol. 14, 2023, doi: 10.1016/j.jpi.2023.100189.

[22] C. Kavitha, V. Mani, S. R. Srividhya, O. I. Khalaf, and C. A. T. Romero, “Early-stage alzheimer’s disease prediction using machine learning models,” *Frontiers in Public Health*, vol. 10, Mar. 2022, doi: 10.3389/fpubh.2022.853294.

[23] F. Z. Benjelloun, A. Oussous, A. Bennani, S. Belfkih, and A. A. Lahcen, “Improving outliers detection in data streams using LiCS and voting,” *Journal of King Saud University-Computer and Information Sciences*, vol. 33, no. 10, pp. 1177–1185, Dec. 2021, doi: 10.1016/j.jksuci.2019.08.003.

[24] N. R. Kolukula, P. N. Pothineni, V. M. K. Chinta, V. G. Boppana, R. P. Kalapala, and S. Duvvi, “Predictive analytics of heart disease presence with feature importance based on machine learning algorithms,” *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 32, no. 2, pp. 1070–1077, Nov. 2023, doi: 10.11591/ijeecs.v32.i2.pp1070-1077.

[25] S. Dekka, K. N. Raju, D. Manendrasai, and M. M. Pallavi, “Utilizing machine learning algorithms for kidney disease prognosis,” *European Journal of Molecular and Clinical Medicine*, vol. 10, no. 1, pp. 2852–2861, 2023.

[26] S. T. Ahmed and S. M. Kadhem, “Alzheimer’s disease prediction using three machine learning methods,” *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 27, no. 3, pp. 1689–1697, Sep. 2022, doi: 10.11591/ijeecs.v27.i3.pp1689-1697.

[27] A. K. Faieq and M. M. Mijwil, “Prediction of heart diseases utilising support vector machine and artificial neural network,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 26, no. 1, pp. 374–380, Apr. 2022, doi: 10.11591/ijeecs.v26.i1.pp374-380.

[28] K. Moulaei, M. Shanbehzadeh, Z. Mohammadi-Taghiabad, and H. Kazemi-Arpanahi, “Comparing machine learning algorithms for predicting COVID-19 mortality,” *BMC Medical Informatics and Decision Making*, vol. 22, no. 1, Jan. 2022, doi: 10.1186/s12911-021-01742-0.

[29] S. Nusinovici *et al.*, “Logistic regression was as good as machine learning for predicting major chronic diseases,” *Journal of Clinical Epidemiology*, vol. 122, pp. 56–69, Jun. 2020, doi: 10.1016/j.jclinepi.2020.03.002.

[30] T. A. Assegie, “Support vector machine and k-nearest neighbor based liver disease classification model,” *Indonesian Journal of electronics, electromedical engineering, and medical informatics*, vol. 3, no. 1, pp. 9–14, Feb. 2021, doi: 10.35882/ijeeemi.v3i1.2.

[31] M. H. Memon, J. Li, A. U. Haq, and M. H. Memon, “Early stage alzheimer’s disease diagnosis method,” in *2019 16th International Computer Conference on Wavelet Active Media Technology and Information Processing, ICCWAMTIP 2019*, Dec. 2019, pp. 222–225, doi: 10.1109/ICCWAMTIP47768.2019.9067689.

[32] E. A. Bayrak, P. Kirci, and T. Ensari, “Comparison of machine learning methods for breast cancer diagnosis,” Apr. 2019, doi: 10.1109/EBBT.2019.8741990.

[33] S. Ganiger and K. M. M. Rajashekharaiah, “Chronic diseases diagnosis using machine learning,” Dec. 2018, doi: 10.1109/ICCSDET.2018.8821235.

[34] Y. Woo, P. T. C. Andres, H. Jeong, and C. Shin, “Classification of diabetic walking through machine learning: Survey targeting senior citizens,” in *3rd International Conference on Artificial Intelligence in Information and Communication, ICAIIC 2021*, Apr. 2021, pp. 435–437, doi: 10.1109/ICAIIC51459.2021.9415250.

[35] N. L. Fitriyani, M. Syafrudin, G. Alfian, and J. Rhee, “HDPM: an effective heart disease prediction model for a clinical decision support system,” *IEEE Access*, vol. 8, pp. 133034–133050, 2020, doi: 10.1109/ACCESS.2020.3010511.

[36] S. F. Weng, J. Reps, J. Kai, J. M. Garibaldi, and N. Qureshi, “Can machine-learning improve cardiovascular risk prediction using routine clinical data?,” *PLoS ONE*, vol. 12, no. 4, Apr. 2017, doi: 10.1371/journal.pone.0174944.

[37] G. Chugh, S. Kumar, and N. Singh, “Survey on machine learning and deep learning applications in breast cancer diagnosis,” *Cognitive Computation*, vol. 13, no. 6, pp. 1451–1470, Jan. 2021, doi: 10.1007/s12559-020-09813-6.

[38] D. N. Reddy, S. M. Bagali, and S. Sadiya, “Machine learning algorithms for detection: a survey and classification,” *Turkish Journal of Computer and Mathematics Education*, vol. 12, no. 10, pp. 3468–3475, 2021, [Online]. Available: https://www.turcomat.org/index.php/turkbilmat/article/view/5025.

[39] T. H. H. Aldhyani, A. S. Alshebami, and M. Y. Alzahrani, “Soft clustering for enhancing the diagnosis of chronic diseases over machine learning algorithms,” *Journal of Healthcare Engineering*, pp. 1–16, Mar. 2020, doi: 10.1155/2020/4984967.

[40] P. Sonar and K. J. Malini, “Diabetes prediction using different machine learning approaches,” in *Proceedings of the 3rd International Conference on Computing Methodologies and Communication, ICCMC 2019*, Mar. 2019, pp. 367–371, doi: 10.1109/ICCMC.2019.8819841.

[41] O. Terrada, B. Cherradi, A. Raihani, and O. Bouattane, “Classification and Prediction of atherosclerosis diseases using machine learning algorithms,” Apr. 2019, doi: 10.1109/ICOA.2019.8727688.

[42] A. Gupta, R. Kumar, H. S. Arora, and B. Raman, “MIFH: a machine intelligence framework for heart disease diagnosis,” *IEEE Access*, vol. 8, pp. 14659–14674, 2020, doi: 10.1109/ACCESS.2019.2962755.

[43] A. Choudhury and D. Gupta, “A survey on medical diagnosis of diabetes using machine learning techniques,” in *Advances in Intelligent Systems and Computing*, vol. 740, Springer Singapore, 2019, pp. 67–78.

[44] K. B. Krishnendu and S. S. Deepa, “A survey on predicting advanced liver fibrosis using different machine learning algorithms,” *International Journal of Scientific Research in Science, Engineering and Technology*, pp. 177–183, Feb. 2020, doi: 10.32628/ijsrset207138.

[45] M. Mallika and K. S. Babu, “Breast cancer prediction using machine learning algorithms,” *International Journal of Science and Research (IJSR)*, vol. 12, no. 10, pp. 1235–1238, Oct. 2023, doi: 10.21275/sr231015173828.

[46] V. S. R. Kumari, S. Veesa, and S. R. Chevala, “Machine learning algorithms to improve the performance metrics of breast cancer diagnosis.” *GRD Journals- Global Research and Development Journal for Engineering*, vol. 6, no. 1, Dec. 2020.

[47] M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang, “Disease prediction by machine learning over big data from healthcare communities,” *IEEE Access*, vol. 5, pp. 8869–8879, 2017, doi: 10.1109/ACCESS.2017.2694446.

[48] 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, doi: 10.1109/ACCESS.2019.2963053.

[49] J. Neelaveni and M. S. G. Devasana, “Alzheimer disease prediction using machine learning algorithms,” in *2020 6th International Conference on Advanced Computing and Communication Systems, ICACCS 2020*, Mar. 2020, pp. 101–104, doi: 10.1109/ICACCS48705.2020.9074248.

[50] S. Shamshirband, M. Fathi, A. Dehzangi, A. T. Chronopoulos, and H. Alinejad-Rokny, “A review on deep learning approaches in healthcare systems: Taxonomies, challenges, and open issues,” *Journal of Biomedical Informatics*, vol. 113, Jan. 2021, doi: 10.1016/j.jbi.2020.103627.

BIOGRAPHIES OF AUTHORS

Aicha Oussous Ph.D. Student in Faculty of Sciences, Mohammed V University in Rabat. Computer Science Engineer, a degree from the National School of Applied Sciences, Agadir, Morocco, in 2013. License diploma in mathematics and Computer Sciences from the Faculty of Sciences, Ibn Zohr University, Agadir, Morocco, in 2010. She can be contacted at email: aicha_oussous@um5.ac.ma.

Abderrahmane Ez-Zahout is currently an Associate Professor at: Department of Computer Sciences/Faculty of Sciences/Mohammed V University. He graduated Ph.D. in Computer Sciences from ENSIAS College of IT. And he is an active member of IPSS Team (intelligent processing systems and security). He can be contacted at email: abderrahmane_ezzahout@um5.ac.ma.

Soumia Ziti is a full professor and researcher at the Faculty of Sciences of Mohammed V University in Rabat since 2007. She obtained her Ph.D. in computer science specializing in graph theory from the University of Orleans in France, along with a diploma in advanced studies in fundamental computer science. She also holds a Baccalaureate in Mathematical Sciences and completed her undergraduate studies in mathematics and physics, specializing in mathematics, at Hassan II University in Morocco. Furthermore, she earned a master’s degree in science and technology in computer science from the same institution. Her research interests encompass a wide range of topics including graph theory, information systems, artificial intelligence, data science, software development, database modelling, big data, cryptography, and numerical methods and simulations. Pr. Ziti has contributed extensively to these fields with over than eighty publications in esteemed international journals and conferences. Additionally, she plays a pivotal role in coordinating, participating or assessing in various educational and socio-economic or research projects. She can be contacted at email: soumia.ziti@fsr.um5.ac.ma.