A Data-Driven University Wellness Program for Cardiovascular Disease Prevention Using Machine Learning-Based Health Risk Classification

Adrales, Lorelyn F.[1], Ruiz, Rue Flora P.[2]

[1]College of Eng'g, Arch & Technology, Notre Dame of Dadiangas University, General Santos City, Philippines

[2]College of Health & Sciences, Notre Dame of Dadiangas University, General Santos City, Philippines

DOI: https://doi.org/10.51244/IJRSI.2026.1306000152

Received: 07 June 2026; Accepted: 12 June 2026; Published: 27 June 2026

ABSTRACT

Annual physical examinations generate valuable employee health data; however, these records are often underutilized for preventive healthcare planning in higher education institutions. This study analyzed the health records of 478 employees of Notre Dame of Dadiangas University (NDDU), General Santos City, using machine learning classification techniques, namely ID3 entropy-based Decision Tree (DT), K-Nearest Neighbor (KNN), and Naïve Bayes (NB). The analysis focused on four cardiovascular disease (CVD) risk markers: cholesterol, creatinine, uric acid, and urinalysis. The findings revealed that 12.6% of employees had elevated cholesterol levels, 13.8% had high uric acid levels, 3.6% showed elevated creatinine levels, and 12.8% had abnormal urinalysis results. These results indicate the presence of cardiovascular and metabolic health risks among a notable portion of the university workforce. While the Decision Tree and Naïve Bayes models achieved perfect classification scores, these results are interpreted with caution due to the pre-classified nature of the dataset and the possibility of data leakage. In contrast, the K-Nearest Neighbor model demonstrated more realistic performance, with F1 scores ranging from 72% to 92% across the different health markers. Based on the identified health risks, the study developed the Marist Workplace Wellness Program for Cardiovascular Disease Prevention. The proposed program includes health education, regular screening, nutrition counseling, stress management, and preventive health interventions tailored to the needs of university employees. The study highlights the value of institutional health records in supporting evidence-based wellness initiatives and demonstrates how machine learning can assist organizations in identifying employee health risks. However, machine learning models should be used only as screening and decision-support tools and not as substitutes for professional medical diagnosis and clinical evaluation.

Keywords: cardiovascular disease (CVD), machine learning, decision tree, university wellness program, employee health

INTRODUCTION

Cardiovascular disease (CVD) remains one of the leading causes of morbidity and premature mortality among working-age adults worldwide. In the Philippines, the Department of Health consistently reports CVD as among the top causes of death, with risk factors including high cholesterol, elevated uric acid, kidney dysfunction, and urinary abnormalities remaining prevalent yet often undetected until clinical complications arise (DOH, 2023). University employees, who represent a stable, health-insured workforce, present an ideal population for preventive health monitoring and intervention.

Notre Dame of Dadiangas University (NDDU) in General Santos City conducts mandatory annual physical examinations for all employees. These examinations generate longitudinal health records encompassing blood chemistry (cholesterol, creatinine, uric acid) and clinical microscopy (urinalysis). Despite the availability of this data, its potential for informing evidence-based wellness programming has not been systematically explored. This gap represents both an institutional risk and a missed opportunity for proactive health management.

The intersection of health informatics and machine learning offers a powerful approach to addressing this gap. Classification algorithms — particularly Decision Trees, Naïve Bayes, and K-Nearest Neighbor (KNN) — have demonstrated strong applicability to binary health data classification tasks in similar institutional contexts (Wang & Zuo, 2022; Ahsan et al., 2022; Reddy et al., 2021). However, most such studies stop at performance evaluation and do not translate their findings into actionable health programming.

This study bridges that gap. Drawing on ML-based classification of NDDU employee health records, it pursues two interrelated objectives: (1) to classify employee CVD risk levels using Decision Tree, KNN, and Naïve Bayes algorithms applied to four CVD markers; and (2) to design a technology-informed Cardiovascular Disease Prevention and Wellness Program grounded in the classification findings. The study contributes to the literature by demonstrating a replicable, institution-level model for translating machine learning health data classification into structured workplace wellness initiatives.

REVIEW OF RELATED LITERATURE

Machine Learning in Employee and Workplace Health

Machine learning has been widely applied to classify and predict health conditions across diverse datasets. Dhillon and Singh (2024) provide a comprehensive survey of ML applications in healthcare, affirming that classification models can reliably identify risk patterns from structured health data. In the workplace context, ML-driven health classification has been used to support occupational health screening, reduce diagnostic delay, and prioritize medical referrals (Kolling et al., 2021).

The Decision Tree algorithm, in particular, is valued in health informatics for its interpretability — the resulting rules can be understood by non-technical health officers without specialized data science training (Ahsan et al., 2022; Nugroho et al., 2016). Naïve Bayes similarly performs well on binary categorical health data, while KNN provides a distance-based perspective that is sensitive to dataset distribution and normalization (Hota & Dewangan, 2016; Xing & Bei, 2019).

CVD Risk Markers: Cholesterol, Creatinine, Uric Acid, and Urinalysis

Elevated total cholesterol (≥200 mg/dL) is a well-established independent risk factor for atherosclerosis, coronary artery disease, and ischemic stroke (WHO, 2023). Creatinine is a reliable biomarker of renal function; elevated levels signal reduced glomerular filtration, which is independently associated with cardiovascular morbidity (MFMER, 2024). Uric acid, when elevated above 7.2 mg/dL, is linked to gout, hypertension, and metabolic syndrome — all recognized CVD precursors (Sampa et al., 2020). Urinalysis abnormalities, including proteinuria and abnormal pH, reflect systemic metabolic disturbances with cardiovascular implications (Fitrani et al., 2019).

The co-occurrence of two or more elevated CVD markers in the same individual substantially increases cardiovascular risk, underscoring the importance of multi-marker classification approaches over single-marker screening (Reddy et al., 2021).

Workplace Wellness Programs and Evidence Base

Workplace wellness programs (WWPs) have been extensively studied as mechanisms for reducing employee health risks, lowering medical costs, and improving productivity. Goetzel et al. (2014) established that well-designed, evidence-based WWPs are correlated with measurably healthier employee outcomes. In a large-scale longitudinal study, Pesis-Katz et al. (2020) demonstrated that personalized wellness programs reduced the 10-year CVD risk score in nearly half of at-risk participants. Song (2019) found that wellness program exposure was associated with significantly higher rates of positive self-reported health behaviors.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VI June 2026

Critically, the evidence base for WWPs is most robust when programs are designed from empirical health data rather than generic population-level guidelines (Howarth et al., 2017). This study responds directly to that imperative by grounding its proposed wellness program in institution-specific machine learning classification results.

Statement of the Problem

This study sought to:

1. Determine the CVD risk profile of NDDU employees based on four markers: cholesterol, uric acid, creatinine, and urinalysis, using machine learning classification.

2.Design and evaluate a Decision Tree model using the ID3 entropy-based induction framework for CVD marker classification and compare its performance with KNN and Naïve Bayes.

3.Translate the ML classification findings into a structured, evidence-based Cardiovascular Disease Prevention and Wellness Program for NDDU employees.

METHODOLOGY

Research Design

The study employed a quantitative-descriptive design combining data mining and machine learning classification. The descriptive component characterized the health risk profile of NDDU employees across four CVD markers. The machine learning component evaluated the predictive performance of three classification algorithms on the same dataset.

Dataset and Data Source

The dataset comprised 478 employee health records collected from the NDDU University Clinic for academic years 2019–2020 and 2022–2023. Records were purposively selected based on the availability of all four required CVD marker results: total cholesterol, uric acid, creatinine, and urinalysis. Each record was de-identified, with employees assigned randomized ID codes. Qualitative data (employee sex: Male = 0, Female = 1) were included as a covariate.

Data collection was conducted with approval from the Office of the Vice President for Administration and adhered to R.A. 10173 (Data Privacy Act of 2012) and the Joint A.O. No. 2016-0002 Privacy Guidelines for the Philippine Health Information Exchange.

Data Preprocessing

Prior to model development, the raw employee health records underwent a series of preprocessing procedures to ensure data quality and suitability for classification. Duplicate entries and data inconsistencies resulting from data encoding were identified and removed. The outcomes of the cardiovascular disease (CVD) markers were then transformed into binary values, where 0 represented normal results and 1 represented high or abnormal results, based on clinically established reference ranges from recognized health authorities such as the World Health Organization (WHO), Mayo Clinic, UCSF Health, and the National Institutes of Health (NIH). For the K-Nearest Neighbor (KNN) algorithm, continuous variables were normalized using Min-Max scaling to prevent features with larger numerical ranges from disproportionately influencing distance calculations. Records containing missing values were excluded from the analysis to maintain dataset integrity, considering the relatively small dataset size and the availability of complete records for classification.

Machine Learning Models

Three machine learning classification algorithms were implemented using Python 3.x through the Anaconda distribution and Jupyter Notebook environment. The Decision Tree algorithm used in this study employs entropy

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VI June 2026

and information gain as the node-splitting criterion — a framework rooted in the ID3 induction algorithm and widely applied in health data classification (Fitrani et al., 2019; Ahsan et al., 2022; Nugroho et al., 2016). Fitrani et al. (2019) specifically implemented the ID3 algorithm through a web-based WEKA environment for health record classification, while Adebayo (2017) demonstrated its application alongside C4.5 for hypertension risk stratification. The entropy values generated in the present study's decision tree structures — ranging from 0.29 to 0.717 at root nodes and resolving to 0.0 at all leaf nodes — are consistent with the behavior of entropy-based induction as documented across these studies. The K-Nearest Neighbor (KNN) model was implemented using the KNeighborsClassifier with a value of k = 5 and the Euclidean distance metric. To ensure accurate distance-based computations, feature normalization was performed before model training. The Naïve Bayes (NB) classifier was implemented using the Gaussian Naïve Bayes algorithm, which is suitable for handling the binary classification structure of the CVD marker outcomes. For model development and evaluation, the dataset was divided into training and testing subsets using the train_test_split() function, with 70% (n = 335) allocated for training and 30% (n = 143) for testing. Stratified sampling was employed to maintain the proportional distribution of Normal and High classifications across both subsets, thereby ensuring balanced representation of outcome classes during model training and evaluation.

Performance Evaluation

Model performance was evaluated using accuracy, precision, recall (sensitivity), and F1 score. Confusion matrices were generated for each model. Performance was assessed on the held-out test set (30%) to approximate generalizability.

RESULTS AND DISCUSSION

Employee CVD Risk Profile

Analysis of the 478 employee health records revealed the following risk distribution across the four CVD markers:

Table 1. CVD Risk Distribution Among NDDU Employees (n=478)

CVD Marker	Total Tested	High / Abnormal (n)	High / Abnormal (%)
Cholesterol	478	**60**	**12.6%**
Uric Acid	478	**66**	**13.8%**
Creatinine	478	**17**	**3.6%**
Urinalysis	478	**61**	**12.8%**

Uric acid emerged as the most prevalent risk marker, with 13.8% of employees classified as high — a finding clinically significant given uric acid's association with gout, hypertension, and metabolic syndrome. Cholesterol (12.6%) and urinalysis abnormalities (12.8%) showed comparable prevalence rates, suggesting that approximately one in eight NDDU employees carries elevated cardiovascular risk on these indicators. Creatinine abnormalities were less prevalent (3.6%), but carry clinical gravity as indicators of renal dysfunction.

Importantly, the presence of elevated markers across multiple categories in the same individual — while not quantified in this cross-sectional analysis — would represent compounded CVD risk. This reinforces the need for multi-marker wellness screening rather than single-indicator approaches.

Decision Tree Structure: What the Trees Actually Reveal

The Decision Tree algorithm produced a notably shallow, two-level structure for each CVD marker. For cholesterol, the root node split on 334 training samples — 274 Normal and 60 High — with initial entropy of

0.679. After a single split, both child nodes achieved entropy = 0.0: 274 Normal left, 60 High right. Equivalent two-level structures emerged for uric acid (268 Normal / 66 High), creatinine (317 Normal / 17 High), and urinalysis (273 Normal / 61 High).

This outcome requires careful interpretation. Each CVD marker was already binary-encoded at the point of input (0 = Normal, 1 = High), meaning classification labels were derived from clinically established thresholds applied during preprocessing — not inferred from raw continuous values. The Decision Tree was effectively trained to re-classify labels that were already determined by rule. A single binary split is the mathematically minimal structure needed to recover this pre-encoded boundary. The resulting trees do not discover complex diagnostic patterns; they recover a predetermined categorical mapping.

This is a critical distinction: the tree's structural simplicity reflects the nature of the input data encoding, not the model's genuine predictive power. Training on raw continuous biomarker values rather than pre-binarized outcomes would present a genuinely complex classification problem, and would constitute a more rigorous test of the model's generalizability.

Comparative Model Performance and Critical Evaluation

Table 2 presents the comparative performance of all three models. These results must be read alongside the structural analysis in Section 5.2 and the methodological transparency note below.

Table 2. Comparative Model Performance Across CVD Markers (30% Test Set, n=143)

CVD Marker	Model	Accuracy	Precision	Recall / Sensitivity	F1 Score
Cholesterol	Decision Tree	100%*	100%*	100%*	100%*
	KNN	87%	80%	87%	83%
	Naïve Bayes	100%*	100%*	100%*	100%*
Creatinine	Decision Tree	100%*	100%*	100%*	100%*
	KNN	94%	89%	94%	92%
	Naïve Bayes	100%*	100%*	100%*	100%*
Uric Acid	Decision Tree	100%*	100%*	100%*	100%*
	KNN	94%	89%	94%	92%
	Naïve Bayes	100%*	100%*	100%*	100%*
Urinalysis	Decision Tree	100%*	100%*	100%*	100%*
	KNN	80%	65%	80%	72%
	Naïve Bayes	100%*	100%*	100%*	100%*

The performance evaluation of the Decision Tree and Naïve Bayes models yielded 100% accuracy, precision, recall, and F1 scores across all four CVD markers, a result that warrants critical methodological scrutiny rather than uncritical acceptance. Three plausible explanations account for this outcome. First, the CVD marker outcomes were binarized during preprocessing using fixed clinical reference thresholds, meaning the Decision Tree was effectively trained to re-learn the same threshold rule already applied to generate the classification labels. This constitutes a trivially separable task in which ceiling-level accuracy is a structural artifact of the data encoding rather than evidence of genuine model performance. Second, if the binarization step was applied to the entire dataset prior to the train-test split, the test set would no longer represent genuinely unseen data — a condition known as data leakage, which artificially inflates all reported performance metrics. Third, the dataset of 478 records drawn from a single institution may lack sufficient distributional variance to expose the model to genuinely ambiguous classification cases, resulting in accuracy estimates that are unlikely to replicate on more diverse or externally sourced data.

It is important to note, however, that these methodological limitations do not undermine the validity of the wellness program developed in this study. The program's foundational evidence base is the descriptive CVD risk prevalence profile derived directly from clinical records and presented in Table 1, which is entirely independent of model performance metrics. The machine learning component is positioned in this study as a screening and triage instrument rather than a clinical diagnostic tool, and any institutional deployment of the classification models must be preceded by rigorous re-validation through k-fold cross-validation and testing on expanded, multi-institution datasets.

Despite these caveats, the comparative results are meaningful. KNN's substantially more variable and moderate performance — ranging from 72% accuracy on urinalysis (F1 = 72%) to 94% on creatinine and uric acid (F1 = 92%) — is itself diagnostic. This variability reflects each marker's distributional characteristics and KNN's sensitivity to class imbalance and feature scaling, factors masked by the ceiling-level scores of DT and NB.

KNN's tendency toward false positives — most pronounced for urinalysis (precision 65%, recall 80%) — directly informs the wellness program design. In a workplace health context, false positives generate unnecessary referrals, employee anxiety, and wasted clinical resources. This finding supports positioning ML classification as a first-pass signal requiring clinical confirmation by a qualified health officer before any intervention is prescribed, rather than as an autonomous diagnostic system.

KNN's poor urinalysis performance (F1 = 72%) is further explained by the composite nature of the urinalysis label, which collapses four sub-parameters (pH, potassium, protein, sodium) into one binary outcome, creating within-class heterogeneity that distance-based algorithms cannot resolve without sub-parameter-level features. Future iterations should treat each urinalysis parameter independently.

Clinical Interpretation by Marker

Cholesterol: The 12.6% prevalence of high cholesterol among NDDU employees is comparable to national prevalence estimates for Filipino working adults (Philippine Heart Association, 2022). The clean DT separation (entropy 0.679 at root, 0.0 at leaves) affirms that cholesterol classification is highly discriminable with the available data, making it a reliable candidate for automated wellness screening.

Uric Acid: The highest prevalence rate (13.8%) and strong KNN performance (94%) on this marker suggest that uric acid distribution among NDDU employees is relatively well-separated between normal and high-risk groups. This finding aligns with Sampa et al. (2020), who found a mean uric acid level of 6.63 mg/dL — borderline high — among a comparable employee sample, underscoring the need for regular monitoring among at-risk individuals.

Creatinine: Though the least prevalent marker (3.6%), elevated creatinine signals early renal impairment and carries disproportionate cardiovascular significance. The sex-stratified reference ranges (men: 0.74–1.35 mg/dL; women: 0.59–1.04 mg/dL) introduce a clinically meaningful dimension that future models should retain as separate features rather than collapsing into a single binary variable.

Urinalysis: The composite nature of urinalysis — aggregating pH, potassium, protein, and sodium into one binary outcome — likely explains KNN's comparatively poor performance (80% accuracy, F1 = 72%). Future iterations should treat each urinalysis sub-parameter as an independent feature to improve model sensitivity and support more granular clinical interpretation.

Ethical Consideration

This study was conducted in accordance with R.A. 10173 (Data Privacy Act of 2012, Philippines) and the Joint A.O. No. 2016-0002 Privacy Guidelines for the Philippine Health Information Exchange. Ethical approval was obtained from the NDDU Office of the Vice President for Administration. All employee health records were de-identified prior to analysis, with randomized ID codes replacing personal identifiers. No individual employee can be identified from any data presented in this study. Findings are reported in aggregate form only.

The authors declare no conflict of interest. This study received no external funding and was conducted as part of the institutional research program of Notre Dame of Dadiangas University.

The Marist Workplace Wellness Program for Cardiovascular Disease Prevention

Program Rationale

The classification results provide a clear empirical basis for targeted wellness programming. With approximately 1 in 8 NDDU employees carrying elevated cholesterol, uric acid, or urinalysis abnormalities, a passive wait-and-treat approach is insufficient. The literature on workplace wellness programs confirms that data-informed, proactive interventions — particularly those anchored in personalized health risk profiles — yield measurable reductions in CVD risk scores (Pesis-Katz et al., 2020; Goetzel et al., 2014).

The program is theoretically grounded in Nola Pender's Health Promotion Model, which posits that health behaviors are shaped by individual characteristics, perceived benefits, interpersonal influences, and situational factors. By designing a structured, supportive workplace environment that addresses these determinants, the program aims to shift employee health behaviors toward sustained CVD risk reduction.

Program Objectives

1. Increase awareness and knowledge of CVD risk factors among all NDDU employees.
2. Enable early identification of at-risk employees through annual ML-assisted health screening.
3. Provide targeted lifestyle modification support for employees with elevated CVD markers.
4. Facilitate medical referral pathways for employees with clinically significant risk levels.
5. Evaluate and report program outcomes annually to support evidence-based program evolution.

Program Components and Implementation Matrix

Marist Workplace Wellness Program For Cvd Prevention — Implementation Matrix

Database	Objectives	Strategies	Target Audience/ Participants	Personal/ Unit Involved	Target Schedule of Implementation	Success Indicator
Among 334 sample, 60 have a high cholesterol	• Organize seminars/ lecture on Cardiovascular Disease Prevention. • Provide informative brochures and materials to educate employees about Cardiovascular Disease and	Education and awareness on Cardiovascular Disease (CVD)	All NDDU Employees	Research and Publication Center (RPC) School Clinic College of Health Sciences (CHS) Faculty and Students	Year 1	All offices provided with brochures

	emergency response to cardiovascular attack. • Promote tobacco free workplace. • Host health fairs to raise awareness about Cardiovascular Disease (CVD) and Prevention					
334 sample, 66 have high uric acid	• Conduct annual health risk assessment of cardiovascular conditions. • Identify the need for disease management. • To provide individualized guidance to employees who need help with their diet particularly those with lipid disorder or other risk factors for chronic disease.	Early Detection and Screening Nutrition Counseling	All NDDU Employees	School Clinic	Year 1	
334, 61 have creatinine	• Provide education about stress reduction and stress management through symposium. • Work with primary care physicians and other health care professionals to refer patients diagnosed with	Stress Management Community Partnerships	All NDDU Employees NDDU Employees diagnosed with Cardiovascular Disease (CVD)	School Clinic/ Nurse	Year 2	

	Cardiovascular Disease (CVD) and integrate Cardiovascular Disease 9CVD) prevention measures into routine care.					

Role of Machine Learning in Program Sustainability

A key innovation of this wellness program is the integration of the Decision Tree classification model as an annual screening tool. At each cycle of the physical examination, employee health data can be re-classified using the trained DT model, generating updated risk profiles that track changes over time. This creates a feedback loop between health intervention and evidence-based program refinement — an approach well-aligned with the precision health paradigm emerging in institutional health management.

The Decision Tree's interpretable IF-THEN rule structure (e.g., 'IF cholesterol ≥ 200 mg/dL THEN classify as High') enables school nurses and health officers without data science backgrounds to understand and apply the model's output directly in clinical decision-making — a critical consideration for sustainability in resource-constrained institutional settings.

CONCLUSION

This study demonstrates that institutional employee health data, when processed through machine learning classification, can serve as a meaningful evidence base for targeted workplace wellness programming — provided that the methodological foundations of the ML component are critically examined and honestly reported. The ID3 entropy-based Decision Tree, alongside Naïve Bayes and KNN, was applied to classify employee CVD risk across four markers.

The descriptive CVD risk profile derived from 478 NDDU employee records is the study's most robust and policy-actionable finding: approximately one in eight employees carries elevated risk on cholesterol, uric acid, or urinalysis indicators, and one in twenty-eight shows elevated creatinine — a clinically significant renal and cardiovascular risk signal. These prevalence figures, derived directly from clinical records, are independent of model performance and constitute a sound empirical basis for the proposed Marist Workplace Wellness Program.

The ML component of the study must be interpreted with appropriate caution. The 100% accuracy scores reported for Decision Tree and Naïve Bayes reflect the structure of the pre-binarized input data rather than genuine generalized predictive power. The models were trained on labels already derived from fixed clinical thresholds, making perfect classification an expected rather than impressive outcome on this dataset. Future validation using raw continuous biomarker values, k-fold cross-validation, and multi-institution data is a prerequisite for any clinical or administrative deployment of these models. KNN's more moderate and variable performance (72–94%) provides a more honest reflection of classification difficulty, and its false-positive tendency directly shapes the program's design principle of using ML outputs as triage signals confirmed by qualified health officers rather than autonomous diagnoses.

The Marist Workplace Wellness Program, grounded in Pender's Health Promotion Model and structured around the institution's own risk data, represents a replicable template for Philippine universities seeking to operationalize their annual physical examination data. Future work should expand the dataset longitudinally and across institutions, disaggregate composite markers, and track year-on-year changes in CVD risk prevalence to evaluate program effectiveness — completing the evidence cycle from data, to classification, to intervention, to measurable health outcomes.

RECOMMENDATIONS

The study recommends the following:

1. Enhance the data set used in this study by increasing the number of test cases and by adding attributes related to employee personal characteristics, diet details and other health data

2. Institutionalize ML-assisted CVD screening as part of the annual physical examination cycle at NDDU, using the Decision Tree model as a first-pass risk classifier.

3. Expand the wellness program to include employees' dependents and student populations, extending the institution's health promotion mandate.

4. Develop a web-based or mobile health dashboard to communicate individual risk classifications to employees in a clear, privacy-compliant, and actionable format.

5. Explore the application of ensemble methods (Random Forest, XGBoost) and deep learning on expanded datasets to further improve classification generalizability.

6. Adapt the Marist wellness program for cardiovascular disease prevention to address the diverse health needs and circumstance of individuals which in a workplace.

7. Implement the Marist Wellness Program for Cardiovascular Disease Prevention to reduce the incidence of Cardiovascular -Disease (CVD) and its related complications.

8. Conduct a follow-up study to assess whether participation in the Marist Workplace Wellness Program measurably reduces CVD risk marker prevalence over three to five years.

REFERENCES

1. Adebayo, P. (2017). Predictive model for the classification of hypertension risk using decision trees algorithm. American Journal of Mathematical and Computer Modelling. https://www.researchgate.net/publication/314284545
2. Ahsan, M., Luna, S., & Siddique, Z. (2022). Machine-learning based disease diagnosis: A comprehensive review. Healthcare, 10, 541. https://doi.org/10.3390/healthcare10030541
3. Dhillon, A., & Singh, A. (2024). Machine learning in healthcare data analysis: A survey. International Online Medical Council (IOMC). https://www.iomcworld.org/articles/machine-learning-in-healthcare-data-analysis-a-survey-44184.html
4. Fitrani, A., Rosid, M., Findawati, Y., Rahmawati, Y., & Anam, A. (2019). Implementation of ID3 algorithm classification using web-based WEKA. Journal of Physics: Conference Series, 1381, 012036. https://doi.org/10.1088/1742-6596/1381/1/012036
5. Goetzel, R. Z., Henke, R. M., Tabrizi, M., Pelletier, K. R., Loeppke, R., Ballard, D. W., & Metz, R. D. (2014). Do workplace health promotion (wellness) programs work? Journal of Occupational and Environmental Medicine, 56(9), 927–934.
6. Hota, H. S., & Dewangan, S. (2016). Classification of health care data using machine learning technique. International Journal of Engineering Science Invention, 5(9), 17–20.
7. Howarth, D., Quesada, J., Silva, J., Judycki, S., & Mills, P. R. (2017). The impact of digital health interventions on health-related outcomes in the workplace: A systematic review. Digital Health, 3. https://doi.org/10.1177/2055207617716113
8. Hu, L., & Li, L. (2022). Using tree-based machine learning for health studies: Literature review and case series. International Journal of Environmental Research and Public Health, 19, 16080. https://doi.org/10.3390/ijerph192316080
9. Jones, D., Molitor, D., & Reif, J. (2020). What do workplace wellness programs do? Evidence from the Illinois workplace wellness study. Quarterly Journal of Economics, 135(3), 1747–1791.

10. Kolling, M., Furstenau, L., Sott, M., Rabaioli, B., Ulmi, P., Bragazzi, N., & Tedesco, L. (2021). Data mining in healthcare: Applying strategic intelligence techniques to depict 25 years of research development. International Journal of Environmental Research and Public Health, 18(6), 3099. https://doi.org/10.3390/ijerph18063099
11. Mayo Foundation for Medical Education and Research (MFMER). (2024). Creatinine test. Mayo Clinic. https://www.mayoclinic.org/tests-procedures/creatinine-test/about/pac-20384646
12. Nugroho, S., Gunawan, D., & Yani, A. (2016). Decision tree induction for classifying the cholesterol levels. Proceedings of the 2nd International Conference on Science, Technology, and Humanity, 231. https://www.researchgate.net/publication/305859333
13. Ong, A. K. S., Kurata, Y. B., Castro, S. a. D., Hazel, D. R. B., & Tomines, A. P. J. (2022). Factors influencing the acceptance of telemedicine in the Philippines. Technology in Society, 70(C). https://doi.org/10.1016/j.techsoc.2022.102040
14. Patterson, P. D., Weaver, M. D., Fabio, A., Studnek, J. R., Renn, M. L., Rosen, C. B., & Hostler, D. (2016). Association between poor sleep, fatigue, and safety outcomes in emergency medical services providers. Prehospital Emergency Care, 16(1), 86–97.
15. Pesis-Katz, I., Williams, G. C., Battaglia, T. A., Wolf, R. L., Chinman, M., & Muennig, P. (2020). Long-term impact of a personalized employee wellness program. American Journal of Health Promotion, 34(2), 134–142.
16. Reddy, V., Meghana, P., Reddy, N., & Rao, B. (2021). Prediction on cardiovascular disease using decision tree and Naïve Bayes classifiers. Journal of Physics: Conference Series, 2161. https://doi.org/10.1088/1742-6596/2161/1/012015
17. Sampa, M. B., Hossain, M. N., Hoque, M. R., Islam, R., Yokota, F., Nishikitani, M., & Ahmed, A. (2020). Blood uric acid prediction with machine learning: Model development and performance comparison. JMIR Medical Informatics, 8(10), e18331. https://doi.org/10.2196/18331
18. Song, Z., & Baicker, K. (2019). Effect of a workplace wellness program on employee health and economic outcomes: A randomized clinical trial. JAMA, 321(15), 1491–1501.
19. Subramani, S., Varshney, N., Anand, M. V., Soudagar, M. M., Al-Keridis, L. A., Upadhyay, T. K., Alshammari, N., Saeed, M., Subramanian, K., Anbarasu, K., & Rohini, K. (2023). Cardiovascular diseases prediction by machine learning incorporation with deep learning. Frontiers in Medicine, 10, 1150933. https://doi.org/10.3389/fmed.2023.1150933
20. University of California San Francisco Health (UCSF). (2024). Uric acid test. https://www.ucsfhealth.org/medical-tests/uric-acid
21. Wang, L., & Zuo, K. (2022). Medical data classification assisted by machine learning strategy. Computational and Mathematical Methods in Medicine, 2022, 9699612. https://doi.org/10.1155/2022/9699612
22. World Health Organization (WHO). (2023). Cardiovascular diseases (CVDs): Key facts. https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)
23. Xing, W., & Bei, Y. (2019). Medical health big data classification based on KNN classification algorithm. IEEE Access. https://doi.org/10.1109/ACCESS.2019.2960130