



The Role of Machine Learning and Metaheuristic Optimization in Enhancing Health Risk Prediction: A Review

Marwa Radwan^{1,*}, Shomona Jacob²

¹Faculty of Artificial Intelligence, Delta University for Science and Technology, Mansoura, Egypt

²School of ICT, Faculty of Engineering, Design and Information & Communications Technology (EDICT), Bahrain Polytechnic, Isa Town, Bahrain

Emails: marwa.radwan@deltauniv.edu.eg; shomona.jacob@polytechnic.bh

Abstract

The present review aims to describe the impact of machine learning techniques in health risk prediction, including the progress, drawbacks, and potential development. ML approaches in health care have become more effective in risk prediction than simple regression techniques because of their accuracy, scalability, and personalization. A Statistical tool like Decision trees, Support vector machines, and neural networks allows or examine non-linear genetic and environmental interactions with lifestyle factors. The review's main points are the increase in relevance of more complex types of models like ANN-PSO, a combination of two algorithms for feature selection, higher prediction accuracy, and other fields, including healthcare. These innovations have shown a unique success rate in identifying diseases, including obesity, diabetes, and any cardiovascular diseases, for better prevention measures and avenues of cure. Nevertheless, there are several difficulties: inferior quality of the data, the question of privacy, and explaining the decision-making of the modern complex models. Solving these issues calls for effective data governance, explainable artificial intelligence, and a multi-disciplinary approach to create and deliver transparency and fairness. As mentioned in the review, feature importance analysis like SHAP also carries plenty of significance for enhancing model interpretability to chase positive alterations. Concerning the outlook, implementing ML in the current HC system will require investment in data platforms, clinician expertise, and broader, suitable systems. As a result, new opportunities for using ML in connection with population health, patient and client outcomes, and receiving individualized care point out the further evolution of the transforming impact of technology. This paper offers an understanding of how health risk prediction and the public health strategy might benefit from new applications of ML and how the moral and practical issues of this new application of the technology may be dealt with.

Keywords: Machine learning; Health risk prediction; Hybrid models; Public health; Data privacy; explainable AI

1. Introduction

Modern leading health threats, including obesity, diabetes, and cardiovascular diseases, pose serious risks for global populations and healthcare systems. Conventional analytical models, primarily evolving from traditional statistical techniques such as regression analysis, have yet to help represent existing correlations between the genetic, environmental, and behavioral causes underlying these human ailments. Thus, one has seen a shift towards applying ML methods to improve health risk prediction accuracy. At the same time, these methods offer flexibility in dealing with copious quantities of data and the ability to find intricate relationships and associations, resulting in improved and individualized risk assessment. This paper seeks to

define the current literature on machine learning approaches to health risk prediction relative to traditional techniques, new sets of algorithms in risk prediction, and implications for improving public health dilemmas.

1.1 Traditional Approaches in Health Risk Prediction

The conventional approaches to health risk assessment differ from the social media risk assessment models. The former has methodological roots in statistical approaches like linear and logistic regression, where simplicity and interpretability are valued most. These approaches have been fundamental in analyzing data involving only two classes and association between variables. However, because of their assumptions, such as linearity and independence of predictors, they have not been instrumental in modeling health risks. For instance, logistic regression models are computationally efficient and work well with clean separated data but cannot manage large-scale complex data common in current Halts settings. However, they rely on data and data quality and thus need help, especially in areas with limited health information [1].

Such restrictions have inspired new sophisticated modeling methods embracing machine learning, which is superior to classical models in deciphering intricate relationships between factors like genes, behaviors, and settings. Consequently, state-of-the-art techniques used today include gradient boosting and regularized logistic regression; feature selection and ensemble learning have improved the predictive capability of such models. Lasso and Ridge regression techniques conquer overfitting by reducing or nullifying the coefficients into a small value that helps generalize unseen data. In Gradient Boosting, an example of a tool is Boost, in which the models create decision trees iteratively by learning from the mistakes of previous steps. Therefore, it is efficient in large-scale, sparse data. Such developments represent a mobile progression from traditionally based statistical frameworks to more complex, data-driven methods of health risk prediction [2].

1.2 Machine Learning in Health Risk Prediction

Machine learning (ML) is the key to risk assessment in analyzing copious amounts of medical and lifestyle data to find correlations contributing to risk calculation. Methods such as decision trees, support vector machines, and neural networks, as well as optimize performance compared to conventional methods in analyzing extensive data for non-linear relationships. For example, some models show the potential of a person developing certain diseases such as diabetes, heart disease, or Alzheimer's illness. Due to their ability to utilize various data ranging from medical records and images to the environment, ML arrives at precise and timely risk predictions and can then recommend ways of preventing complications and appropriately customizing patients' treatment plans [3].

Recent works have proven the potential of using a higher ML level in real-life healthcare scenarios. One such case is the UCSF AI model, which can predict candidates likely to develop Alzheimer's disease up to seven years before its onset using clinical data only. To this extent, the role of ML in making meaningful information from regular clinical records must be highlighted. Likewise, the MIT-developed Mirai model employs mammogram raw images after integrating the patients' risk factors in screening BC up to five years ahead. Such developments prove ML's versatility across different clinical settings and its potential for enhancing diagnostic performance in various population groups to provide vital equity in health besides boosting early intervention gains [4].

1.3 Hybrid Machine Learning Models and Innovations

Machine learning models that combine several techniques signify a new level of development of predictive analysis tools to increase the effectiveness of their application. Of these innovations, it is possible to mention the Artificial Neural Network-Particle Swarm Optimization (ANN-PSO) model. This model utilizes the optimization properties of a Particle Swarm Optimization (PSO) technique in combination with the unique learning capabilities of neural networks, making it easier to fine-tune the model and choose the right features. For instance, studies have been elucidating its use in the risk of obesity, which is better than the traditional regression analysis. Such advancement fosters how hybrid models can contribute to efficient and accurate predictions of various aspects of healthcare and, most importantly, how constructive collaboration between different algorithms can enhance the general solution to complicated tasks [5].

All these hybrid techniques are also being extended to other disciplines, such as cardiovascular risk evaluation and mental health assessment. Machine learning algorithms combined with optimization methods were reported in the literature as providing powerful tools for capturing complex, non-linear relationships between variables. For example, in cardiovascular disease studies, socio-environmental parameters

integrated models have better prediction outcomes in factors such as heart failure and rehospitalization than the clinical parameters-only models. Such mixed strategies' growing use supports the premise that they could redefine the applicability of predictive health care models by providing much more refined insights into the disease risk probabilities and intervention plans, as are evident from the latest developments in graph neural networks and collaborative filtering of health records [6].

1.4 Impact of Machine Learning on Public Health Interventions

One of the revolutionary innovations affecting the public health approach is machine learning (ML), which is used to produce more accurate and nuanced approaches. Using ML models, patient health information, such as genetic makeup, environment, and lifestyle, is incorporated into developing and delivering prevention and treatment measures. For instance, the SHAP algorithm describes risks for diabetes and obesity and can be used to fashion interventions and policies. Such extraction capability helps detect pathophysiological changes early, intervene at the right time, and avoid complications [7].

Furthermore, ML is critically important in monitoring public health, with the features of predicting the occurrence of cases of spread of infectious diseases amongst others by analyzing various other sources of information, including but not limited to social media, weather, and geographical information. Such systems improve early detection systems and allow prompt intervention to contain the events. Such applications highlight the significance of ML in contributing to improved healthcare systems when it comes to disease prevention and outbreak control and in contributing to the initiative-taking delivery of healthcare services [8].

1.5 Challenges and Future Directions

In risk prediction for health risks, hurdles that present it with a narrower application hinder the application of machine learning (ML). The first challenge is obtaining the right quality and variety of datasets. Most ML models require patient data, including genetic information, behavior change, and environment. However, it can be challenging to obtain such data as may be restricted by laws, privacy, and the unstipulated healthcare industry. In addition, the bias inherent in specific datasets may compromise model precision and equitable performance in populations and settings where such systems are implemented. These problems illustrate the importance of better data acquisition and data anonymization methods for ensuring fairness representation in health datasets and refining the efficacy of the models obtained.

The second critical issue is related to known weaknesses inherent in the specificity and oversights of ML models. While utilizing methods such as deep learning, high accuracy is attainable; however, for healthcare practitioners, models tend to be “black boxes.” Such lack of transparency raises ethical and clinically pertinent issues, especially in key life decisions requiring explanation. As a next step, future work should focus more on creating model-interpretor pairs based on off-the-shelf algorithms and enhancing general end-to-end XAI frameworks to guarantee the output of both high accuracy of the model and the possibility of utilizing the individual predictions. However, important challenges remain in integrating HCML with current healthcare platforms and educating clinicians to adopt new techniques to acquire the most out of HL.

Consequently, the use of machine learning leads to significant advances in the methods of health risk prediction in the aspects of accuracy, process scalability, and client specificity. Integrating algorithms like the ANN-PSO model offers a better prospect of improving these prediction models. As machine learning advances, applying this capability within public health programs may increase the efficacy of healthcare interventions and the target approach. However, data quality and model interpretability are key factors that must be addressed to ensure these advances can be implemented in real-life healthcare settings. Public health is likely to benefit from the efforts of machine learning in dealing with complex problems in the modern world.

2. Literature Review

The rise of different ailments, such as obesity, has created awareness that warrants the development of better prediction models for improved interactive social health interventions. Like the regression models, traditional models are extremely limited in more common diseases with genetic, environmental, and behavioral components. With new machine learning (ML) methods, health risk prediction's accuracy and reliability will soon be enhanced. Ultimate ML models, which consist of the ANN-PSO model, present more quantitative and complex analysis since they incorporate multiple data types and utilize enhanced algorithms—applying

machine learning in healthcare results in incredibly positive outcomes in providing individualized prevention and treatment strategies for key public health issues.

Accurate crude oil price forecasting is critical for economic activities such as energy trading, risk management, and investment planning, as highlighted in the study [9]. While deep learning models have proven effective for forecasting, their accuracy still needs to be improved by challenges such as hyperparameter tuning and sensitivity to external factors like global events. To overcome these issues, the research proposes a hybrid method combining mathematical optimization methods with an ensemble of five neural network architectures. Using the Grey Wolf Optimizer (GWO) across four stages—feature selection, data preparation, model training, and forecast blending, the approach demonstrated improved performance in forecasting three-ahead days of Brent crude oil prices. The results yielded a mean squared error (MSE) of 0.000127, indicating enhanced accuracy when benchmarked against traditional models.

Efficient speech-emotion recognition (SER) is critical for developing intuitive human-computer interaction systems, as outlined in [10]. The research addresses challenges in achieving high accuracy through optimized feature extraction, selection, and model tuning. It introduces a machine-learning framework that employs the Hybrid Bat Algorithm (HBA) to optimize a Random Forest model, termed the Bat Random Forest Hybrid Metaheuristic Algorithm (BRHAMO). The study evaluates spectral and hybrid speech features using speech corpora such as RAVDESS, SAVEE, and the ANAKE Hindi corpus. The BRHAMO model achieves superior accuracy rates of 93.8%, 85.4%, and 89.8% in the hybrid feature category for these datasets, outperforming benchmark models like vanilla Random Forest, Gradient Boost, Adaptive Boost, and Support Vector Machines in accuracy, precision, recall, and F1 score.

Advances in early breast cancer detection and improved treatments have significantly increased survival rates, as detailed in [11]. While traditional screening methods such as mammography, MRI, and biopsies are effective, they often involve prohibitive costs and risks. Thermal imaging has emerged as a safer alternative but is limited in adoption due to its reliance on detecting temperature changes. The research introduces a novel interpretable computer-aided diagnosis (CAD) system leveraging Explainable Artificial Intelligence (XAI). It employs hybrid optimization algorithms, Hybrid Particle Swarm Optimization (HPSO) and Hybrid Spider Monkey Optimization (HSMO) to balance accuracy, feature selection, and hyperparameter tuning. Using handcrafted feature extraction methods like LBP, HOG, and Gabor Filters combined with the Shapley Additive Explanations (SHAP) framework, the study achieves high performance on the DMR-IR dataset. With HSMO, the model reached an accuracy of 98.27% and an F1-score of 98.15%, selecting only 25.78% of HOG features, demonstrating improved performance and enhanced interpretability.

Load forecasting is a critical aspect of the power industry, requiring techniques that minimize prediction errors to ensure uninterrupted power supply and significant cost savings, as discussed in [12]. The research presents a comprehensive framework for short-term load forecasting using metaheuristic algorithms organized into three layers: data decomposition, forecasting, and Optimization. The data-decomposition layer extracts key features from input data, while the forecasting layer employs statistical and machine learning models to predict demand. The optimization layer fine-tunes model parameters using metaheuristic algorithms to enhance accuracy and stability. Hybrid models integrating these layers address challenges such as local minimum and premature convergence, achieving better prediction performance. The study includes a systematic review of metaheuristic algorithms and deep learning methods, provides a taxonomy of related works, and identifies future research directions for hyperparameter tuning in forecasting models.

In the research presented in [13], agriculture's critical role in the Indian economy is emphasized, particularly the challenges in crop recommendation and yield prediction due to variables such as soil type, climatic conditions, and irrigation methods. The study proposes the Gorilla Troops Optimization with Deep Learning-based Crop Recommendation and Yield Prediction Model (GTODL-CRYPM) to address these issues. This model integrates Gorilla Troops Optimization (GTO) with Long Short-Term Memory (LSTM) networks for efficient crop recommendation, optimizing LSTM parameters using GTO. A Deep Belief Network (DBN) also makes accurate crop yield predictions. Extensive experiments conducted on the Crop Recommendation Dataset and Crop Yield Prediction Dataset demonstrate the GTODL-CRYPM model's superior performance, achieving a maximum accuracy of 99.88% and an R^2 score of 99.14%, outperforming other methods.

In the research presented in [14], a novel sign language recognition system is developed to bridge the communication gap between individuals with hearing and speech disabilities and those without such challenges. The system leverages the recently devised metaheuristic WAR Strategy optimization algorithm

to address the trade-off between accuracy and computing time in sign language recognition. Spatial and temporal features are extracted using Linear Discriminant Analysis (LDA) and the Gray-Level Co-occurrence Matrix (GLCM) methods. The WAR Strategy algorithm is employed for feature optimization and hyperparameter tuning of six standard machine learning models, enhancing recognition efficiency. Evaluated on datasets covering American, Arabic, and Malaysian sign languages, the system achieved recognition accuracies ranging from 93.11% to 100% with training times of 0.038 to 10.48 seconds. Experimental results highlight the system's superior accuracy, generalization, computational efficiency, and complexity.

Paper [15] has established a new Mountain Gazelle Optimization with Deep Ensemble Learning-based Intrusion Detection (MGODEL-ID) technique to improve the security aspects of Smart Grids (SGs). Recently, SGs have integrated smart meters and renewable energy sources into their network while needing IDS to protect against cyber risks. It also adopts unique features for data preprocessing where the MGODEL-ID model uses z-score normalization while feature selection in this model uses the MGO model. Intrusion detection is performed using a set of classifiers based on LSTM, DAE, and ELM, with the DBO algorithm involved for hyperparameter optimization. Simulation outcomes indicate that the MGODEL-ID model can better detect intrusions than other models, thereby strengthening the smart grid security against cyber-attacks.

In the research presented in [16], a modified convolutional neural network (CNN) architecture, U-Net, was optimized using Particle Swarm Optimization (PSO) to enhance brain tumor segmentation from magnetic resonance imaging (MRI) data. To achieve notable performance improvements, the method focuses on hyperparameter tuning, particularly learning and dropout rates. The integration of skip connections and dropout layers effectively captures intricate features. It mitigates overfitting, achieving Dice Similarity Coefficient (DSC) and Jaccard Index (JI) values of up to 94.14% and 89.02%, respectively, across tumor types such as Meningioma, Glioma, and Pituitary. Comparative analysis highlights the approach's superiority over conventional methods, demonstrating its potential in advancing automated tumor segmentation in clinical settings.

In the study referenced as [17], a novel wrapper feature selection method combining Orthogonal Learning (OL) and a rime optimization algorithm (RIME) was introduced, termed MRIME. The MRIME algorithm optimizes feature selection (FS) without compromising classifier performance, particularly in bladder cancer (BC) diagnosis. This method was integrated with Support Vector Machine (SVM) for classification, creating the MRIME-SVM model. Evaluated across eight BC datasets, MRIME-SVM outperformed other metaheuristic algorithms regarding classification accuracy. Additionally, MRIME demonstrated its competitiveness in global optimization tasks through comparative analysis with various optimization algorithms, further solidifying its potential in advancing BC diagnostic accuracy.

In the research presented in [18], the rapid evolution of cyberattacks on the Internet of Things (IoT) necessitates novel approaches to address zero-day attacks. Traditional Intrusion Detection Systems (IDS) often struggle to detect unidentified attacks and raise concerns about data privacy and security. A Binary Snake Optimizer with Deep Learning-Enabled Zero-Day Attack Detection and Classification (BSODL-ZDADC) method is proposed to address these challenges. This approach combines metaheuristic techniques with deep learning for improved attack recognition and classification. The BSODL-ZDADC method uses Z-score normalization, the Binary Snake Optimizer (BSO) for feature selection, and an attention-based bidirectional gated recurrent unit (ABi-GRU) for attack detection. Experimental results on the TON-IoT dataset show that the BSODL-ZDADC method achieves superior accuracy, reaching 98.28% compared to other models.

In the study referenced as [19], lung diseases, primarily caused by air pollution, smoking, and respiratory infections, pose significant health risks and diminish quality of life. The study explores the use of audio file analysis for detecting lung diseases, with data derived from 236 audio files. The audio features were extracted using the VGG16 pre-trained transfer-learning model, resulting in 4096 features, which were then classified using Support Vector Machines (SVM), achieving an accuracy of 92.40%. Hyperparameter optimization was conducted with the Slap Swarm Algorithm (SSA) to enhance performance, boosting accuracy to 96.61%. The results demonstrate that combining VGG16 for feature extraction and SSA for Optimization offers an effective method for detecting lung diseases, highlighting the potential of integrating transfer learning and methodology in improving medical diagnostics.

In the research presented in [20], antibiotics in water sources pose significant risks to both environmental and public health. The study focused on using machine learning (ML) techniques to develop models that predict the adsorption capacity of ciprofloxacin (CIP) antibiotic from contaminated water. A total of ten ML algorithms were evaluated using various performance metrics, including R2, MSE, MAE, and others, to assess their robustness. The hyperparameters of these models were optimized using Bayesian Optimization. The results revealed that the His Gradient Boosting (HGB) model provided the best performance, with an MAE of 0.1865 and an R2 value of 0.9999. The optimized model predicted the highest antibiotic adsorption (99.28%) under specific conditions, highlighting the potential of combining advanced ML algorithms with nano adsorbents to mitigate antibiotic pollution in water sources.

The study [21] identified [21] the Internet of Vehicles (IOV) as a crucial IoT application for traffic management, road safety, and driving comfort. The paper proposes an intrusion detection system based on Genetic Algorithm (GA) and Machine Learning (ML) to enhance IOV security. The performance of various ML algorithms was optimized using GA, with the random forest model achieving the highest detection accuracy of 99.64%. The study suggests that future research will focus on developing more effective ML models for other IoT applications.

In the study referenced as [22], brain tumor diagnosis is highlighted as a critical task for prognosis and planning treatment for patients with brain cancer. The paper introduces a novel approach for diagnosing brain tumors in MRI scans using deep learning, specifically focusing on Residual/Shuffle Networks. A modified metaheuristic algorithm, Augmented Falcon Finch Optimization (AFFO), is proposed to enhance the network's performance. AFFO utilizes bio-inspired principles to optimize hyperparameters, improving the reliability and accuracy of the deep learning model. The method's performance is evaluated on a standard MRI dataset and compared to existing techniques, such as RESNET, ALEXNET, VGG-16, Inception V3, and U-Net, demonstrating the effectiveness of combining Residual/Shuffle Networks with AFFO for brain tumor classification.

In the study referenced as [23], the growth of machine learning (ML) approaches has led to innovations in hydraulic fracturing design. The research presents a physics-based dataset consisting of sixty-two parameters used to construct an ML model. The dataset was evaluated using a transfer learning approach to enhance predictive capabilities with an actual field dataset. Three training-testing combinations were used: synthetic-synthetic, synthetic-real, and real real, to demonstrate the effectiveness of transfer learning. Neural networks were applied to maximize production results, combined with hyperparameter optimization routines and a particle swarm optimization loop. The transfer learning approach showed enhanced performance, with improvements of 15.12% in root mean square error (RMSE) and 15.88% in mean absolute percentage error (MAPE) compared to the pure data-trained model. The optimized parameter space resulted in production values 14.2% higher than the initial predictions, with 88% of instances being optimal compared to actual values. The findings highlight the potential of physics-based ML and transfer learning for improving predictive accuracy and optimizing outcomes in hydraulic fracturing and similar fields with limited data.

In the study referenced as [24], the accurate prediction of solar energy generation is crucial for efficient energy management in Internet of Things (IoT) devices. Current forecasting models often need help to address the dynamic nature of weather conditions, and traditional methods typically exhibit limited accuracy and scalability. To address these challenges, the authors propose the SEP-CVSGCNN-IoT method, which utilizes a Complex-Valued Spatio-Temporal Graph Convolutional Neural Network (CVSGCNN) optimized for solar energy prediction. The method begins by collecting data from solar panels and weather forecasts, which is preprocessed using Data-Adaptive Gaussian Average Filtering (DAGAF) to eliminate unwanted data and replace missing values. The optimal features are selected using the Nutcracker Optimization (NCO) algorithm, and the enhanced data is input into the CVSGCNN for solar energy prediction. The Dipper Throated Optimization Algorithm (DTOA) is also employed to optimize the CVSGCNN classifier's weight parameters. The results demonstrate that the SEP-CVSGCNN-IoT method achieves 18.46%, 26.34%, 15.69%, and 20.84% higher accuracy and 18.24%, 23.77%, 24.34%, and 16.29% lower mean absolute error compared to existing techniques, including deep learning enhanced solar energy prediction and AI-driven IoT methods.

In the research presented in [25], sign language is the primary communication method for individuals with hearing and speech disabilities. However, comprehension of sign language by those without disabilities remains a significant challenge, resulting in communication disparities. Despite various effective machine-

learning techniques, a trade-off exists between accuracy and computation time in sign language recognition systems. The authors propose a novel sign language recognition system developed using the recently introduced metaheuristic WAR Strategy optimization algorithm to address this. The system extracts spatial and temporal features through Linear Discriminant Analysis (LDA) and Gray-Level Co-occurrence Matrix (GLCM) methods. The WAR Strategy optimization algorithm is applied in two phases: first, to optimize the extracted feature set and second, to fine-tune the hyperparameters of six standard machine-learning models to achieve precise and efficient recognition. The system was evaluated on datasets from multiple languages (American, Arabic, and Malaysian) containing various sign language variations. The results show that the proposed system achieves recognition accuracy ranging from 93.11% to 100% with training times between 0.038 and 10.48 seconds. Experimental outcomes demonstrate that the proposed system excels in efficiency, time, complexity, generalization, and accuracy.

In the study referenced as [26], antibiotics in water sources present significant risks to public health and the environment, necessitating effective monitoring and intervention strategies. The research investigates machine learning (ML) techniques to develop models for analyzing the adsorption capacity of ciprofloxacin (CIP) from contaminated water. Ten ML algorithms were assessed based on various performance metrics, including the Coefficient of Determination (R²), Mean Square Error (MSE), and Root Mean Square Error (RMSE). Hyperparameter optimization was performed using the Bayesian optimization algorithm, and feature importance analysis was conducted to evaluate the significance of operational variables. Among the models evaluated, the Hist Gradient Boosting (HGB) model demonstrated the best performance, with an MAE of 0.1865 and an R² value of 0.9999. The optimized model predicted the highest antibiotic adsorption (99.28%) under conditions of 10 mg/L CIP, 357 mg/L CuWO₄@TiO₂ adsorbent, 60 minutes of contact time, room temperature, and a near-neutral pH of 7.5. The findings highlight the potential of combining advanced ML algorithms and nano adsorbents to address the growing issue of antibiotic contamination in water sources.

In the research presented in [27], efficient speech-emotion recognition (SER) is a key challenge in developing natural and intuitive human-computer interaction systems. The study optimizes feature extraction, selection, and model tuning to enhance SER accuracy. The proposed machine-learning framework, based on metaheuristic principles, includes two main components: feature extraction and selection. It introduces an optimized random forest model, the Bat Random Forest Hybrid Meta-Heuristic Algorithm (BRHAMO), which utilizes the hybrid bat algorithm (HBA) for hyperparameter tuning. The framework was evaluated using three speech corpora—RAVDESS, SAVEE, and the novel ANAKE Hindi speech corpus. The results showed that the BRHAMO model achieved an accuracy of 81%, 79.6%, and 77.6% for the RAVDESS, SAVEE, and ANAKE datasets, respectively, using spectral features. When hybrid features were applied, the model improved to 93.8%, 85.4%, and 89.8% accuracy, respectively. Comparative analysis with benchmark machine learning models, including vanilla Random Forest, gradient boost, adaptive boost, and support vector machines, demonstrated that the metaheuristic approach outperformed individual classifiers regarding accuracy, precision, F1 score, and recall across both feature categories.

As documented in [28], obesity is increasing in prevalence worldwide; therefore, more efficient prediction methods are required. In particular, the traditional regression models may not effectively identify the intricate relationships between genetics, environment and behavior contributing to obesity. In other words, this research seeks to determine the feasibility of using machine learning in improving obesity risk prediction. Afterward, the authors perform data preprocessing and evaluation before applying several supervised learning algorithms, the ANN-PSO hybrid model of which is new. Compared with the regression approach, the proposed ANN-PSO model analyzed reached an accuracy of 92%. Moreover, the importance of features was assessed using SHAP, which provides more sophisticated information about the features that affect obesities, such as genetic factors. The study also underlines how emerging levels of machine learning will help improve healthcare delivery through a precision medicine approach and address the problem of obesity.

The analysis of the results presented in Table 1 indicates that a broad range of methodologies has been employed in addressing a wide array of problems in healthcare, energy, agri-food, and cybersecurity fields, typically utilizing metaheuristic optimization algorithms as a component of a machine learning system. These methods have shown enhancements in the prediction of accuracy, identification of features, and randomization of the models, resulting in solutions improvements. Researchers continuously focus on applying machine learning and intelligent Optimization to open new areas of application of innovative technologies that can help enrich the life experience of different industries.

Table 1: Summary of Literature Review

Study	Research Area	Proposed Method	Key Findings
[9]	Crude Oil Price Forecasting	Hybrid Metaheuristic Optimization with Ensemble of Neural Networks	Improved accuracy and reduced MSE compared to traditional models.
[10]	Speech Emotion Recognition	Hybrid Bat Algorithm (HBA) with Random Forest	Enhanced accuracy and outperformed benchmark models.
[11]	Breast Cancer Detection	Hybrid Optimization Algorithms (HPSO and HSMO) with Explainable AI	Improved accuracy and interpretability.
[12]	Short-Term Load Forecasting	Metaheuristic Algorithms with Data Decomposition and Forecasting Layers	Improved accuracy and stability.
[13]	Crop Recommendation and Yield Prediction	Gorilla Troops Optimization with Deep Learning (GTODL-CRYPM)	High accuracy and R ² score.
[14]	Sign Language Recognition	WAR Strategy Optimization with Machine Learning	High accuracy and efficiency.
[15]	Intrusion Detection in Smart Grids	Mountain Gazelle Optimization with Deep Ensemble Learning (MODEL-ID)	Improved intrusion detection accuracy.
[16]	Brain Tumor Segmentation	Particle Swarm Optimization with U-Net	Enhanced segmentation accuracy.
[17]	Bladder Cancer Diagnosis	Orthogonal Learning and RIME Optimization with SVM (MRIME-SVM)	Improved classification accuracy.
[18]	Zero-Day Attack Detection in IoT	Binary Snake Optimizer with Deep Learning (BSODL-ZDADC)	Improved attack detection and classification accuracy.
[19]	Lung Disease Detection	VGG16 and Slap Swarm Algorithm (SSA)	Enhanced lung disease detection accuracy.
[20]	Antibiotic Adsorption Prediction	Machine Learning with Bayesian Optimization	Improved prediction accuracy and identified optimal conditions.
[21]	Intrusion Detection in Internet of Vehicles	Genetic Algorithm and Machine Learning	Enhanced intrusion detection accuracy.
[22]	Brain Tumor Diagnosis	Augmented Falcon Finch Optimization with Residual/Shuffle Networks	Improved brain tumor classification accuracy.

[23]	Hydraulic Fracturing Design	Transfer Learning with Neural Networks and Hyperparameter Optimization	Improved predictive accuracy and Optimization.
[24]	Solar Energy Prediction in IoT	Complex-valued spatial-temporal Graph Convolutional Neural Network (CVSGCNN) with Optimization	Improved prediction accuracy and reduced error.
[25]	Sign Language Recognition	WAR Strategy Optimization with Machine Learning	Improved accuracy and efficiency.
[26]	Antibiotic Adsorption Prediction	Machine Learning with Bayesian Optimization	Improved prediction accuracy and identified optimal conditions.
[27]	Speech Emotion Recognition	Hybrid Bat Algorithm with Random Forest	Enhanced emotion recognition accuracy.
[28]	Obesity Risk Prediction	ANN-PSO Hybrid Model	Improved prediction accuracy and identified influential factors.

In conclusion, the analyzed literature focuses on the increasing importance of machine learning in developing innovative approaches to diseases, such as obesity, from the perspective of public health practice. These superior models augment the predictive capability and explain which factors lead to unhealthy scenarios. Such techniques as SHAP analysis also enhance the ability of these models to display the relative significance of different risks, which makes the model incredibly useful to healthcare practitioners. As healthcare progresses, the application of ML, especially in risk prediction and management, offers an opportunity to enhance individual outcomes and widely held health plans. Finally, the future of public health will rest in advancing machines, learning to solve some of the current health problems.

3. Discussion

The discussion part of the paper focuses on the application of ML to health risk assessment and intervention within the HC context, together with the discussion of the challenges and future perspectives in this field. The transformation from statistical techniques towards modern ML techniques affords a dramatic leap in the exactness and data capacities that have been transformed and progressively enhanced healthcare equipment in the future. An increasing number of hybrid models, including the ANN-PSO, clearly signal the possibility of integrating several algorithms to achieve the best prediction results in different areas of health. However, problems like data quality, privacy, and a better understanding of advanced ML techniques remain the primary concerns. This section ends by discussing future profitable trajectories such as ideals for data access, explainable AI, and operationalization of ML for optimum contributions to the overall healthcare system.

3.1 The Role of Machine Learning in Health Risk Prediction

The use of ML has had a significant impact on health risk prediction because it provides more accurate and efficient approaches than statistical instruments. Indeed, the conventional approaches to health risk assessment, based on logistic regression, among others, although essential in the past, need to capture the complexity and nonlinearity associated with contemporary health information. The application of newer models of ML-like decision trees, support vector machines, and deep learning models has enabled scientists and healthcare givers to capture the multifactorial interactions of genes, environment, and behaviors well. They allow a person-centered risk evaluation, which results in appropriate and timely protective actions [29].

For instance, algorithms like XGBOOST and Lasso regression have shown a way to improve the effectiveness of predictions, eradicating some issues that affect traditional regression models, such as overfitting and multicollinearity. Further, machine learning applications in the prediction of diseases, as well as

the identification of diseases like Alzheimer's or cardiovascular diseases based on patient records, are highly showing the role of AI in preventable health care. Applying these ML models to big data from multiple sources provides better interpretability and, hopefully, better Read: By using big data sets from various sources, these models can make more precise and better clinical decisions and, therefore, better treatment outcomes [30].

3.2 Hybrid Models and Their Impact on Healthcare

Indeed, significant progress in developing hybrid models in such popular machine learning methodologies is a big leap in applying predictive health analytics. These models stand to use a fragment of various algorithms in a way that captures the best from each of them. A key innovation is the ANN-PSO (Artificial Neural Network-Particle Swarm Optimization) model; this combines the possibilities of optimizing the use of Particle Swarm Optimization with the possibilities of Artificial Neural Network's learning competence. This constructive collaboration improves function discrimination, model optimization, and, more generally, prediction, particularly in cases of obesity and cardiological risks. This is why hybrid models are especially popular in cases when linear relations between variables can be assumed, which single-algorithm models fail to recognize, and that is why they are so valuable for the numerous aspects of the health field.

Furthermore, hybrids continue to diversify to new fields as they apply socio-environmental features to clinical ones to enhance the risk factor estimation of cardiovascular diseases. These models have shown enhanced risk prediction capability by incorporating more variables than earlier, presenting differential risks and, thereby, differential interventions. Such advancements propose that hybrid models remain relevant and will progress in future healthcare applications by providing more accurate and specific risk assessments, better health consequences, and better systems characteristics [31].

3.3 Challenges in Data Quality and Interpretability

However, there are still some crucial issues related to improving data quality for health risk prediction using ML and the problems of interpretability of the results of working models. Broad and accurate databases are critical for improving the ML models; however, the problems with privacy and legislation do not allow obtaining such information. That is, there is often no consistency in the data formats, and the data can contain biases that harm the performance of the systems, especially in different contexts like healthcare. For instance, ML models should be developed with diverse data inputs, including clinical, genomics, geo-epidemiological, and lifestyle data. However, obtaining and incorporating such data in a way that respects patient rights and provides fairness to the models is still incredibly challenging [32].

Moreover, due to the sophistication of the algorithms, especially the deep learning ones, there are questions about whether the models make sense. Although such models yield good accuracy, they are most times considered opaque decision-makers; this is because the models work by what is referred to in deep learning as feeding data into the model, and the model spits out the prediction while it is difficult for the healthcare professional to determine how the model reached such a verdict. This lack of transparency calls for ethical and clinical questions whenever such models are adopted in critical situations. Consequently, any future studies should involve the usage of explainable artificial intelligence (XAI) systems, which help increase the effectiveness of predictions while providing information about the background of the prediction. In doing so, the concepts within an ML system can become benign and acceptable for practical application in clinical practice [33].

3.4 Future Directions and Integration into Healthcare Systems

In the future, adding artificial intelligence into the existing care structures is expected to bring smoother recognition of the patient's health risks. An emerging area of focus is deploying ML models within already established healthcare systems. To move these technologies toward their ideal use cases, clinicians need to know how to employ these models correctly, and the models must be smoothly integrated into current practices. In addition, there is an increasing concern about developing international cooperation to increase health data availability and quality, which will be crucial for fine-tuning and testing ML algorithms in the future. As with any new technological application in healthcare, standardization of data harvesting protocols, data protection rules, and data integration requirements will determine the success of applications based on machine learning.

Furthermore, in the future years, with the growing emergence of machine learning research, more emphasis should be placed on creating fair models. Halting the lack of diversification evident in modern ML models will be a significant step towards ensuring that all the population groups receive proper healthcare, minimizing resulting inequalities. Finally, further studies on hybrid models, interpretability, and data availability seem to be the most critical issues for defining the further evolution of public health and medical care in the coming decades [34].

Currently, there are vast opportunities for the application of machine learning in the prognosis of health risks, which, in the context of their further improvement, provides density, individuality, and scalability compared to statistical models. ANN-PSO is one of the many hybrid models that offer a clear line of development to enhance the accuracy of predictions in various healthcare areas. However, to further unlock such possibilities in healthcare, matters including data quality, severe privacy, and model interpretability must be solved. These models are well geared to change public health interventions and clinic outcomes significantly, and they address the increasing complexities that define healthcare systems across the globe due to the incorporation of technology and machine learning into healthcare systems.

4. Conclusion

Machine learning (ML) has taken the field of health risk prediction by storm due to its accuracy, scalability, and capability to model at an individual level, which statistical methods cannot do. Hence, by incorporating enhanced algorithms such as neural networks, decision trees, or even the ANN-PSO, the possibility of ML identifying non-linear relationships in regions of multiple health determinants has been evident. These have led to better diagnosis of ailments like obesity, diabetes, and cardiovascular diseases at an early stage, thus allowing doctors to plan and treat them appropriately. In recent years, the enhanced interpretability of those models using tools like SHAP for feature importance analysis has only helped to fill the gap between complex technologies in the healthcare sector and their everyday application.

Both uses make hybrid ML models a significant advancement in healthcare analytics. Such models combine the best properties of various algorithms, improve the accuracy of the prediction, and extend the range of health risk assessments. For instance, the ANN-PSO model has enhanced accuracy without compromising on problems of feature selection and tuning of the parameters. The use of these models in various health sectors underscores their capacity to enhance the effectiveness of health interventions and health practice in general-oriented clinical practice situations. As such, by integrating computational processes with domain-exploitative knowledge, future possibilities of hybrid models in enhancing healthcare structures are enabled.

Nonetheless, the integration of ML into healthcare has its difficulties, too. As such, challenges like the quality of data, its availability, and data privacy continue to be pressing questions. Further, the black-box nature of most advanced ML techniques raises substantial ethical and practical issues in clinical practice. Mitigating these challenges involves creating good data governance foundations, finding a common way of data collection, and using XAI to foster trust in artificial intelligence. Such challenges require collective endeavors involving researchers, clinicians, and policymakers to address providing adequate and equal health care using ML-driven health innovations.

Towards the future, all upcoming challenges in the ML domain in healthcare will be adaptively resolved, and the technology will be incorporated into the existing healthcare system. Much of the emphasis must turn to building models that are as efficient, fair, and applicable to a wide range of patient populations as effective for accurately making diagnoses. Some key funding areas include data structures, training for healthcare professionals, and leveraging the development of system interfaces to fulfill the most important goals of ML. Suppose innovative ideas are developed, and more joint projects are launched. In that case, ML can become a driving force in healthcare systems for effectively predicting diseases, optimizing patients' conditions, and solving fundamental problems in public healthcare worldwide.

Reference

- [1] Maciej Serda et al., “Synteza i aktywność biologiczna nowych analogów tiosemikarbazonowych chelatorów żelaza,” *Uniwersytet śląski*, vol. 7, no. 1, pp. 343–354, 2013, doi: 10.2/JQUERY.MIN.JS.
- [2] P. Gupta, B. Ding, C. Guan, and D. Ding, “Generative AI: A systematic review using topic modelling techniques,” *Data Inf Manag*, vol. 8, no. 2, p. 100066, Jun. 2024, doi: 10.1016/J.DIM.2024.100066.
- [3] S. E. Abhadiomhen, E. O. Nzeakor, and K. Oyibo, “Health Risk Assessment Using Machine Learning: Systematic Review,” *Electronics (Switzerland)*, vol. 13, no. 22, p. 4405, Nov. 2024, doi: 10.3390/ELECTRONICS13224405/S1.
- [4] A. S. Tang et al., “Leveraging electronic health records and knowledge networks for Alzheimer’s disease prediction and sex-specific biological insights,” *Nature Aging* 2024 4:3, vol. 4, no. 3, pp. 379–395, Feb. 2024, doi: 10.1038/s43587-024-00573-8.
- [5] Y. Zhao, E. P. Wood, N. Mirin, S. H. Cook, and R. Chunara, “Social Determinants in Machine Learning Cardiovascular Disease Prediction Models: A Systematic Review,” *Am J Prev Med*, vol. 61, no. 4, pp. 596–605, Oct. 2021, doi: 10.1016/J.AMEPRE.2021.04.016.
- [6] R. van den Berg, T. N. Kipf, and M. Welling, “Graph Convolutional Matrix Completion,” Jun. 2017, Accessed: Dec. 05, 2024. [Online]. Available: <https://arxiv.org/abs/1706.02263v2>
- [7] S. H. Jacobson and J. A. Jokela, “Artificial Intelligence and Public Health,” *Artificial Intelligence for Healthcare*, pp. 3–12, Apr. 2022, doi: 10.1017/9781108872188.003.
- [8] A. P. Zhao et al., “AI for science: Predicting infectious diseases,” *Journal of Safety Science and Resilience*, vol. 5, no. 2, pp. 130–146, Jun. 2024, doi: 10.1016/J.JNLSSR.2024.02.002.
- [9] M. Alruqimi and L. Di Persio, “Multistep Brent oil price forecasting with a multi-aspect metaheuristic optimization and ensemble deep learning model,” *Energy Informatics* 2024 7:1, vol. 7, no. 1, pp. 1–19, Nov. 2024, doi: 10.1186/S42162-024-00421-4.
- [10] A. Agrawal and A. Jain, “Brhamo: metaheuristic optimization algorithm for speech emotion recognition using spectral and hybrid features,” *Evol Intell*, vol. 18, no. 1, pp. 1–20, Feb. 2025, doi: 10.1007/S12065-024-00994-9/METRICS.
- [11] H. Dihmani, A. Bousselham, and O. Bouattane, “A New Computer-Aided Diagnosis System for Breast Cancer Detection from Thermograms Using Metaheuristic Algorithms and Explainable AI,” *Algorithms* 2024, Vol. 17, Page 462, vol. 17, no. 10, p. 462, Oct. 2024, doi: 10.3390/A17100462.
- [12] U. Mumtahina, S. Alahakoon, and P. Wolfs, “Hyperparameter Tuning of Load-Forecasting Models Using Metaheuristic Optimization Algorithms—A Systematic Review,” *Mathematics* 2024, Vol. 12, Page 3353, vol. 12, no. 21, p. 3353, Oct. 2024, doi: 10.3390/MATH12213353.
- [13] A. Punitha and V. Geetha, “Gorilla troops optimization with deep learning based crop recommendation and yield prediction,” *International Journal of Cognitive Computing in Engineering*, vol. 5, pp. 494–504, Jan. 2024, doi: 10.1016/J.IJCCE.2024.09.006.
- [14] S. T. Abd Al-Latief, S. Yussof, A. Ahmad, S. M. Khadim, and R. A. Abdulhasan, “Instant Sign Language Recognition by WAR Strategy Algorithm Based Tuned Machine Learning,” *International Journal of Networked and Distributed Computing*, vol. 12, no. 2, pp. 344–361, Dec. 2024, doi: 10.1007/S44227-024-00039-8/FIGURES/12.
- [15] S. A. Sharaf et al., “Advanced mathematical modeling of mitigating security threats in smart grids through deep ensemble model,” *Scientific Reports* 2024 14:1, vol. 14, no. 1, pp. 1–22, Oct. 2024, doi: 10.1038/s41598-024-74733-6.
- [16] S. Saifullah and R. Drezewski, “Automatic Brain Tumor Segmentation Using Convolutional Neural Networks: U-Net Framework with PSO-Tuned Hyperparameters,” pp. 333–351, 2024, doi: 10.1007/978-3-031-70071-2_21/TABLES/4.
- [17] M. E. Hosney, E. H. Houssein, M. R. Saad, N. A. Samee, M. M. Jamjoom, and M. M. Emam, “Efficient bladder cancer diagnosis using an improved RIME algorithm with Orthogonal Learning,” *Comput Biol Med*, vol. 182, p. 109175, Nov. 2024, doi: 10.1016/J.COMPBIOMED.2024.109175.
- [18] A. S. Almuflih et al., “Securing IoT devices with zero day intrusion detection system using binary snake optimization and attention based bidirectional gated recurrent classifier,” *Scientific Reports* 2024 14:1, vol. 14, no. 1, pp. 1–22, Nov. 2024, doi: 10.1038/s41598-024-80255-y.

- [19] A. Ciran, S. Ertem, and E. Ozbay, "Optimization-Based Hyperparameter Selection in Deep Learning Methods for Detection of Lung Diseases," 8th International Artificial Intelligence and Data Processing Symposium, IDAP 2024, 2024, doi: 10.1109/IDAP64064.2024.10710803.
- [20] Y. Ahmed et al., "A novel interpretable machine learning and metaheuristic-based protocol to predict and optimize ciprofloxacin antibiotic adsorption with nano-adsorbent," *J Environ Manage*, vol. 370, p. 122614, Nov. 2024, doi: 10.1016/J.JENVMAN.2024.122614.
- [21] M. F. Gul and H. Bakir, "Improving Attack Detection in IoV Systems using GA-based Hyperparameter Optimization," 8th International Artificial Intelligence and Data Processing Symposium, IDAP 2024, 2024, doi: 10.1109/IDAP64064.2024.10711086.
- [22] X. Guo, T. Liu, and Q. Chi, "Brain tumor diagnosis in MRI scans images using Residual/Shuffle Network optimized by augmented Falcon Finch optimization," *Scientific Reports* 2024 14:1, vol. 14, no. 1, pp. 1–20, Nov. 2024, doi: 10.1038/s41598-024-77523-2.
- [23] A. M. Khan, E. Ugarte, A. BinZiad, and A. Alsubaii, "Physics-Informed Machine Learning for Hydraulic Fracturing—Part II: The Transfer Learning Experiment," *ADIPEC*, Nov. 2024, doi: 10.2118/222903-MS.
- [24] A. B. Kathole, D. Jadhav, K. N. Vhatkar, S. Amol, and N. Gandhewar, "Solar energy prediction in IoT system based optimized complex-valued spatio-temporal graph convolutional neural network," *Knowl Based Syst*, vol. 304, p. 112400, Nov. 2024, doi: 10.1016/J.KNOSYS.2024.112400.
- [25] S. T. Abd Al-Latief, S. Yussof, A. Ahmad, S. M. Khadim, and R. A. Abdulhasan, "Instant Sign Language Recognition by WAR Strategy Algorithm Based Tuned Machine Learning," *International Journal of Networked and Distributed Computing*, vol. 12, no. 2, pp. 344–361, Dec. 2024, doi: 10.1007/S44227-024-00039-8/FIGURES/12.
- [26] Y. Ahmed et al., "A novel interpretable machine learning and metaheuristic-based protocol to predict and optimize ciprofloxacin antibiotic adsorption with nano-adsorbent," *J Environ Manage*, vol. 370, p. 122614, Nov. 2024, doi: 10.1016/J.JENVMAN.2024.122614.
- [27] A. Agrawal and A. Jain, "Brhamo: metaheuristic optimization algorithm for speech emotion recognition using spectral and hybrid features," *Evol Intell*, vol. 18, no. 1, pp. 1–20, Feb. 2025, doi: 10.1007/S12065-024-00994-9/METRICS.
- [28] Z. Helforoush and H. Sayyad, "Prediction and classification of obesity risk based on a hybrid metaheuristic machine learning approach," *Front Big Data*, vol. 7, p. 1469981, Sep. 2024, doi: 10.3389/FDATA.2024.1469981/BIBTEX.
- [29] M. Kavitha, G. Gnaneswar, R. Dinesh, Y. R. Sai, and R. S. Suraj, "Heart Disease Prediction using Hybrid machine Learning Model," *Proceedings of the 6th International Conference on Inventive Computation Technologies, ICICT 2021*, pp. 1329–1333, Jan. 2021, doi: 10.1109/ICICT50816.2021.9358597.
- [30] F. Li, Y. Chen, and H. Xu, "Coronary heart disease prediction based on hybrid deep learning," *Review of Scientific Instruments*, vol. 95, no. 1, Jan. 2024, doi: 10.1063/5.0172368/3147153.
- [31] M. González-Del-Hoyo and X. Rossello, "Challenges and promises of machine learning-based risk prediction modelling in cardiovascular disease," *Eur Heart J Acute Cardiovasc Care*, vol. 10, no. 8, pp. 866–868, Oct. 2021, doi: 10.1093/EHJACC/ZUAB074.
- [32] M. Kavitha, G. Gnaneswar, R. Dinesh, Y. R. Sai, and R. S. Suraj, "Heart Disease Prediction using Hybrid machine Learning Model," *Proceedings of the 6th International Conference on Inventive Computation Technologies, ICICT 2021*, pp. 1329–1333, Jan. 2021, doi: 10.1109/ICICT50816.2021.9358597.
- [33] M. González-Del-Hoyo and X. Rossello, "Challenges and promises of machine learning-based risk prediction modelling in cardiovascular disease," *Eur Heart J Acute Cardiovasc Care*, vol. 10, no. 8, pp. 866–868, Oct. 2021, doi: 10.1093/EHJACC/ZUAB074.
- [34] M. Kavitha, G. Gnaneswar, R. Dinesh, Y. R. Sai, and R. S. Suraj, "Heart Disease Prediction using Hybrid machine Learning Model," *Proceedings of the 6th International Conference on Inventive Computation Technologies, ICICT 2021*, pp. 1329–1333, Jan. 2021, doi: 10.1109/ICICT50816.2021.9358597.