



Integrated deep learning for cardiovascular risk assessment and diagnosis: An evolutionary mating algorithm-enhanced CNN-LSTM[☆]



Ahmed Mohammed Ahmed Alsarori, Mohd Herwan Sulaiman*

Faculty of Electrical & Electronics Engineering Technology, Universiti Malaysia Pahang Al-Sultan Abdullah (UMPSA), 26600 Pekan Pahang, Malaysia

ARTICLE INFO

Method name:

Dual Output CNN-LSTM-EMA

Keywords:

Cardiovascular disease
Machine learning
Deep learning
Hybrid model
Convolutional neural network
Long Short-Term Memory
Evolutionary mating algorithm

ABSTRACT

Cardiovascular diseases (CVD) remain the leading cause of mortality worldwide, emphasizing the urgent need for accurate and efficient predictive models. This study proposes a dual-output deep learning model based on a hybrid Convolutional Neural Network–Long Short-Term Memory (CNN-LSTM) model, optimized using the Evolutionary Mating Algorithm (EMA). The model predicts both a continuous risk score and a binary diagnostic outcome, supporting both quantitative assessment and early clinical decision-making. EMA was applied for hyperparameter optimization, demonstrating improved convergence and generalization over conventional methods. Performance was benchmarked against CNN-LSTM models optimized using Particle Swarm Optimization (PSO) and Barnacle Mating Optimization (BMO). The EMA-based model achieved superior results, with a Mean Absolute Error (MAE) of 0.018, Mean Squared Error (MSE) of 0.0006, Root Mean Squared Error (RMSE) of 0.024, and a coefficient of determination (R^2) of 0.98 for risk prediction. For the diagnostic task, the model attained 70 % accuracy and 80 % precision. These findings validate EMA's effectiveness in tuning dual-output deep learning models and highlight its potential in enhancing cardiovascular risk stratification and early diagnosis in clinical settings.

- Dual-output CNN-LSTM model optimized using EMA.
- Continuous risk scores and binary diagnostic classification predictions.
- EMA outperformed PSO and BMO in predictive accuracy and model robustness.

Specifications table

Subject area:	Engineering
More specific subject area:	Machine Learning and Deep Learning
Name of your method:	Dual Output CNN-LSTM-EMA
Name and reference of original method:	
Resource availability:	Cardiovascular dataset, [online] https://www.kaggle.com/datasets/sulianova/cardiovascular-disease-dataset

Background

Chronic diseases represent a profound and escalating global health challenge, accounting for approximately 74 % of all deaths worldwide. In 2023 alone, non-communicable diseases (NCDs) were responsible for over 41 million deaths, underscoring their pervasive impact across populations [1]. In the United States, nearly 88 % of individuals aged 65 and above are affected by at least one

[☆] **Related research article:** None.

* Corresponding author.

E-mail address: herwan@umpsa.edu.my (M.H. Sulaiman).

<https://doi.org/10.1016/j.mex.2025.103466>

Received 16 May 2025; Accepted 24 June 2025

Available online 27 June 2025

2215-0161/© 2025 Published by Elsevier B.V. This is an open access article under the CC BY license

(<http://creativecommons.org/licenses/by/4.0/>)

Table 1
Dataset variable.

Feature	Variable	Missing Values
ID	id	Non
Age	Age	Non
Height	Height	Non
Weight	Weight	Non
Gender	Gender	Non
Systolic blood pressure	ap_hi	Non
Diastolic blood pressure	ap_lo	Non
Cholesterol	Cholesterol	Non
Glucose	Gluc	Non
Smoking	Smoke	Non
Alcohol intake	Alco	Non
Physical activity	Active	Non
Cardiovascular disease	Cardio	Non

chronic condition, emphasizing the critical need for timely diagnosis, continuous monitoring, and personalized intervention strategies [2]. Among the most burdensome of these conditions are cardiovascular diseases (CVD), diabetes, cancer, and stroke [3]. Although chronic diseases are often preventable or manageable, their effective control hinges on early detection and accurate risk stratification [4].

Recent advancements in Artificial Intelligence (AI), particularly in machine learning (ML), have shown promise in improving disease prediction and clinical decision support systems [5–10]. Deep learning—a subset of ML—has demonstrated exceptional performance in capturing complex, nonlinear relationships in medical data. Deep Neural Networks (DNNs), with their hierarchical architecture, are capable of learning abstract representations from high-dimensional data [11–14]. Within this framework, Convolutional Neural Networks (CNNs) are proficient at extracting spatial features, while Long Short-Term Memory (LSTM) networks excel at modeling temporal dependencies in sequential data. The integration of CNN and LSTM layers into a unified architecture enables robust modeling of both spatial and temporal dynamics in patient health records, making it particularly suitable for CVD prediction tasks [15–18].

Despite their effectiveness, deep learning models are highly sensitive to hyperparameter configurations, which can significantly influence predictive performance [19]. Conventional optimization techniques—such as grid search and random search—are often computationally intensive and inefficient in navigating large, multimodal search spaces [20]. In response to these challenges, recent research has turned to bio-inspired metaheuristic algorithms for hyperparameter tuning.

This study investigates the Evolutionary Mating Algorithm (EMA) [21–26], a novel recombination-based evolutionary technique designed to enhance diversity in the search process. EMA introduces dynamic mating strategies to avoid local optima and promote broader exploration of the solution landscape. By integrating EMA with a hybrid CNN-LSTM architecture, we propose a dual-output deep learning framework capable of simultaneously predicting (i) a continuous risk score for cardiovascular disease and (ii) a binary diagnosis classification. This dual-task approach not only enables fine-grained risk assessment but also supports early diagnostic decision-making.

To benchmark its effectiveness, the proposed CNN-LSTM-EMA model is compared against alternative optimization methods, including Particle Swarm Optimization (PSO) and Barnacle Mating Optimization (BMO). The goal is to demonstrate the superiority of EMA in enhancing both predictive accuracy and generalizability for dual-output cardiovascular disease modeling.

Method details

This paper utilizes a cardiovascular disease dataset and all relevant information with regards to the dataset and how it would be utilized could be read in this chapter. Table 2 provides the data configuration details used for training and testing.

Data collection

It is to be noted that the risk assessment would be a continuous variable, while the early diagnosis would be a binary variable. The risk assessment variable is a percentile variable meaning that it could range from 0 – 100 theoretically, while the early diagnosis variable would be a binary variable predicting and/or indicating the presence of a cardiovascular disease. All of these are to be used at the discretion of the medical practitioner. The utilization of these assessments should be exercised with the careful judgment and discretion of a medical professional. It is imperative that a qualified healthcare practitioner assess and interpret the information provided by these models, considering their expertise and the unique circumstances of each case.

The dataset utilized for this research comprises 70,000 patient records, acquired from Kaggle, a trusted and open source for datasets, and its link is in the reference [27] along with their corresponding features outlined in the provided Table 1.

Table 1 exhibits the 13 initial features or variables within the dataset before undergoing preprocessing. While most of these features are self-explanatory, it is crucial to emphasize that the feature ID, which is linked to patient identification, will not be employed in any model throughout this study. Furthermore, to utilize the data from this dataset for the deep learning algorithm proposed in this

paper, it needs to be normalized and transposed before providing it to the models. Normalizing and transposing would be conducted after preprocessing the data and introducing any feature-engineered features/variables.

Exploratory data

The data processing steps were initiated by conducting a thorough examination for missing or duplicate data. To maintain data integrity, no feature selection methods such as PCA or LASSO regression were applied. The dataset was kept unchanged in its original form, ensuring that all features were preserved for deep learning modeling. However, preprocessing steps were applied to handle missing values, outliers, and noise:

- **Missing Values:** The data was searched and checked for any missing data, but nothing was missing.
- **Outliers:** Extreme outliers, detected using the interquartile range (IQR) method, where values for “ap_hi”, “ap_lo”, “weight”, and “height” fell outside the 2.5 % to 97.5 % range, were systematically eliminated.
- **Noise Handling:** Data smoothing techniques, including moving averages, were applied where necessary to mitigate inconsistencies.

By incorporating the segmentation of continuous input into discrete groups or bins, the algorithm gained the capability to discern subtle variations among different data classes. An illustrative example is the age variable, initially provided in days, which was subsequently converted to the default age format (in years) [28–30].

Feature engineering

Beyond addressing outliers, a comprehensive understanding of the dataset used in this paper was gathered, encompassing the type of data and complete dataset statistics. The introduction of new variables was undertaken for utilization in the paper, with MAP and BMI identified as crucial variables in all risk assessment models while the risk variable served as the target for these models.

In the field of medicine, mean arterial pressure (MAP) refers to the average blood pressure experienced by an individual over a single cardiac cycle. This feature serves as a crucial indicator of both peripheral resistance and cardiac output and is expressed as follows:

$$\text{Mean Arterial Pressure (map)} = \frac{2 \text{ ap_lo} + \text{ ap_hi}}{3} \quad (1)$$

Cardiovascular Disease (CVD) risk is notably elevated among those classified as overweight or obese. The classification of weight could be acquired using Body Mass Index (BMI) and is expressed as follows:

$$\text{Body Mass Index (bmi)} = \frac{\text{weight}}{\text{height}^2} \quad (2)$$

The unique aspect of this study lies in its focus on the risk variable, aiming to evaluate the likelihood of a patient developing CVD [31–34]. The rationale behind selecting these components stems from the recognition, supported by previous medical studies [2], that age and cholesterol play pivotal roles in identifying CVD. This variable comprises all the elements such as “Age”, “MAP”, “cholesterol”, and “BMI”, as well as all the other features/elements in the dataset that would be used in the model and could be expressed as follows:

$$\text{Risk} = \frac{\sum \text{Elements in a single observation}}{\sum \text{Highest possible values of each element in the dataset}} \times 100 \quad (3)$$

Table 2 shows all the variables that would be used in this study after analysis and preprocessing have been completed.

Figs. 1 and 2 show all the inputs and outputs before transposing them. It clearly shows the distribution of data within the dataset. It shows which of the variables would be used as inputs and how they are provided as well as how the output would be, in terms of what prediction values it would produce.

Hyperparameters optimization

To accomplish the goal of this paper, 4 hyperparameters of the CNN-LSTM (Convolutional neural network and long short-term memory network) model are optimized by means of utilizing **Evolutionary Mating Algorithm** to lower the Mean Absolute Error (MAE). These four parameters, as shown in Table 3, are the number of filters, filter size, LSTM units, and dropout rate parameters. Acceptable ranges have been provided based on relevant information.

MAE serves as the objective function for enhancing optimization results, which is expressed as follows:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \bar{y}_i| \quad (4)$$

where;

n – stands for the number of observations

y_i – stands for actual values

\bar{y}_i – stands for optimized values

Table 2
Input and Output Variables.

Feature	Variable	Missing Values	Preview	Input/Output
Age	Age	Non	55, 76, 62, 42	Input
Body Mass Index	BMI	Non	1 – Normal 2 – Overweight 3 – Obese	Input
Mean Arterial Pressure	MAP	Non	1 – Normal 2 –Elevated 3 – Elevated (within early stage 1 hypertension) 4 – Stage 1 Hypertension 5 – Stage 2 hypertension	Input
Gender	Gender	Non	1 – Male 2 - Female	Input
Cholesterol	Cholesterol	Non	1 - normal 2 - high 3 - very high	Input
Glucose	Gluc	Non	1 - normal 2 – high 3 – very high	Input
Smoking	Smoke	Non	0 – non-smoker 1 - smoker	Input
Alcohol intake	Alco	Non	0 – doesn't drink 1 - drinks	Input
Physical activity	Active	Non	0 – not active 1 - active	Input
Risk factor	Risk	Non	64.32, 88.15, 72.98 (percentiles)	Output
Early diagnosis	Cardio	Non	0 – not diagnosed 1 - diagnosed	Output

Histograms of Input Data Features

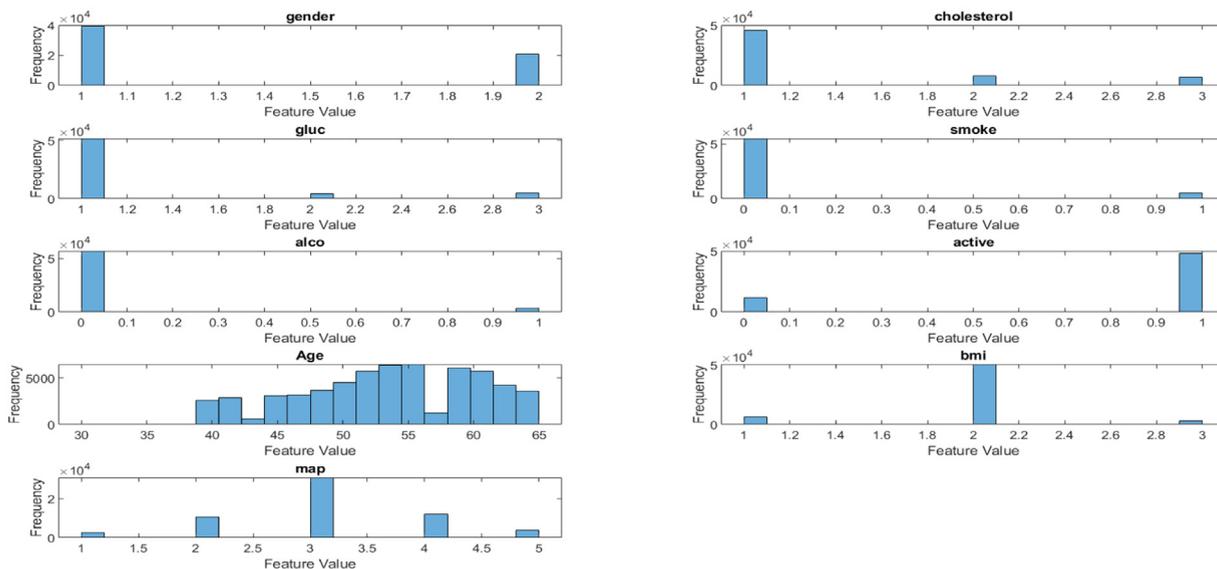


Fig. 1. Input Variables.

This paper utilizes the Evolutionary Mating Algorithm (EMA) as a method to optimize and adjust the hyperparameters of the proposed CNN-LSTM model, which in the context of this paper, would be the base deep learning model as shown in Fig. 3. This optimization is carried out by embedding the EMA function into the CNN-LSTM framework and automating hyperparameter tuning to select the best ones to be used.

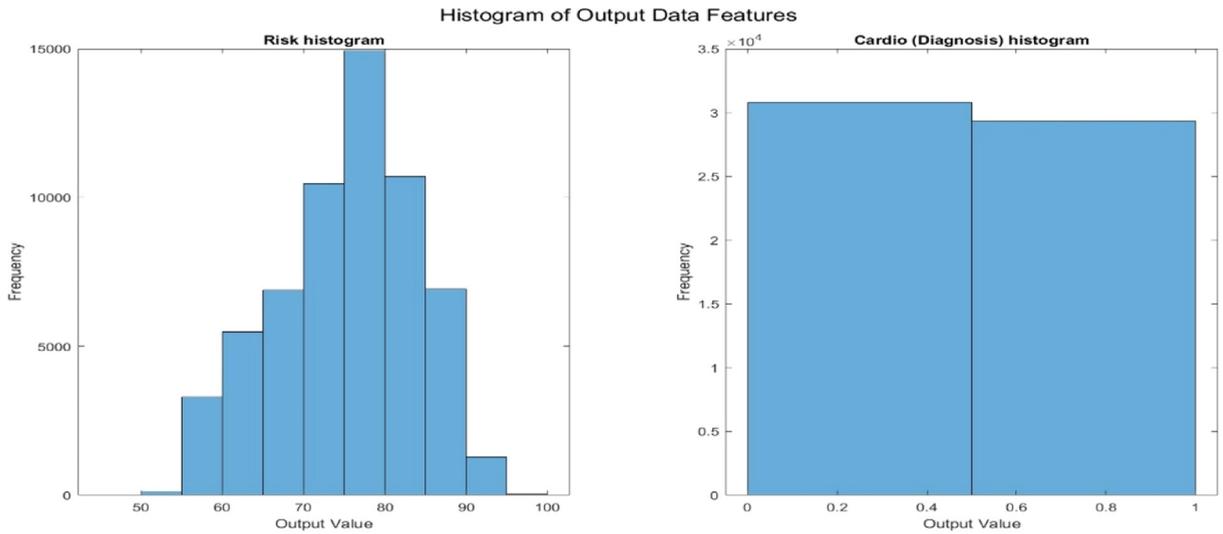


Fig. 2. Output Variables.

Table 3
Hyperparameters.

Parameters	Range	Details
Number of filters	(8–64)	Number of filters in the convolutional layer (lower than 8 leads to poor learning while higher than 64 might risk overfitting)
Filter size	(2–5)	Kernel sizes (size of filters) in the convolutional layer (small kernels, 2–3, might be good for sequential data and learning fine-grained data while 4–5 learns more contextual information)
LSTM units	(32–128)	Number of units in the LSTM layer, the memory capacity to learn long-term dependencies (the larger and complex the dataset, the more units would be needed. Lower than 32 risks than learning the long-term dependencies efficiently while higher than 128 risks overfitting and increased training time)
Dropout rate	(0.1–0.5)	Dropout rate for regularization. (lower than 0.1 leads to overfitting while higher than 0.5 may cause underfitting)

EMA combines elements of genetic algorithms and natural mating processes to function. It is a recently established metaheuristic algorithm that stems from Hardy-Weinberg (HW) principles and is inspired by the mating process in organisms.

$$x_m = \begin{bmatrix} x_1^1 & \dots & x_1^d \\ \vdots & \ddots & \vdots \\ x_{n/2}^1 & \dots & x_{n/2}^d \end{bmatrix} \tag{5}$$

$$x_f = \begin{bmatrix} x_{\frac{n}{2}+1}^1 & \dots & x_{\frac{n}{2}+1}^d \\ \vdots & \ddots & \vdots \\ x_n^1 & \dots & x_n^d \end{bmatrix} \tag{6}$$

$$I_{mates} = 1 + \left[var(x_{m,*}^T) - var(x_{f,*}^T) \right] \tag{7}$$

$$X_{child}^T = \begin{cases} p \cdot X_{m,*}^T + q \cdot X_{f,*}^T & \text{for } I_{mates} \geq 0 \\ p \cdot X_{f,*}^T + q \cdot X_{m,*}^T & \text{for } I_{mates} < 0 \end{cases} \tag{8}$$

$$p = randn(1, d) \tag{9}$$

$$q = (1 - p) \tag{10}$$

This algorithm starts by initializing a population of possible solutions. This is generally the data available. Based on Eq. (5), which is at the initialization stage, two groups are observed. These two groups are acquired after the candidate solution X is split into two as observed from Eqs. (5) and (6): with Xm representing the males' group while Xf represents the females, n referring to the size of the population, and d the dimension of the equation\problem. After initialization is completed, the fitness function is next. It would be evaluated for each member of the population and based on the results acquired, the best solutions from both Xm and Xf would be identified and recorded. Fitter solutions would be favored.

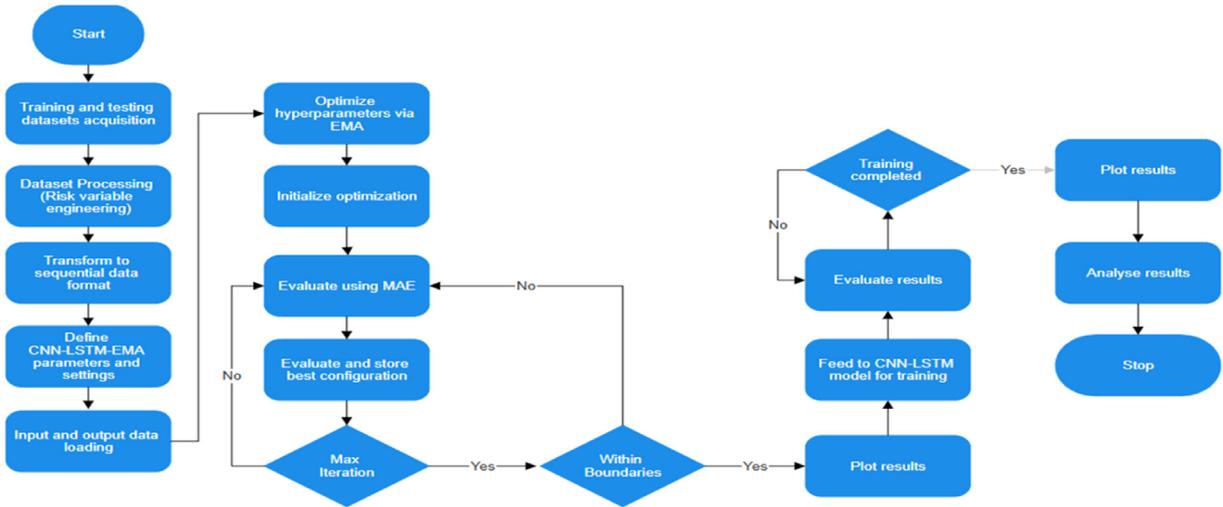


Fig. 3. CNN-LSTM-EMA flowchart.

A crossover would then be performed to produce an offspring solution. This step is the combination of the two selected parent solutions to produce a new solution, the offspring solution. This could be seen in Eq. (7) where I_{mates} is the mating process in EMA based on the opportunity for sexual selection and the $var(x_{m,*}^T)$ and $var(x_{f,*}^T)$ are the variance of the selected male and female to be mated, respectively at iteration T.

Once the mating process concludes, the next process from Eq. (8), the production of new offsprings (X_{child}^T) starts. Here, p refers to normal random distribution, which could further be expressed in Eq. (9) while q could be expressed in Eq. (10). The resulting offspring solutions would then be evaluated based on the fitness functions and the best solutions would be kept, then introduced as parent solutions for the next batch. This process would continue to be repeated for several generations (cycles) until convergence. Furthermore, two parameters would need to be tuned to consider the effects the environment might have, such as predators. These two parameters are Cr and r. Of these two parameters, Cr refers to the probability of crossovers while r refers to predator encountering probability. With environmental changes, the best solution’s characteristics would significantly be altered, since the offspring could be assumed to be dead or alive.

It utilizes the evolutionary mating algorithm to train, evaluate, and improve parameters by means of evolving it to the most suitable ones by running it until the maximum number of iterations has been completed. Hyperparameter tuning was performed using an **adaptive search mechanism**, evaluating model performance iteratively through validation loss reduction. A **cross-validation** approach was not ultimately used due to computational constraints, even though it had realistically the best results; instead, a **60–40 train-test split** was determined as the most effective based on preliminary experiments comparing 70–30, and 80–20 splits.

Proposed model (CNN-LSTM)

After the optimization is completed, it would then be used to train the deep learning model proposed for this paper from scratch (using the training dataset). It would then be evaluated using unseen data from the testing dataset.

Convolutional Neural Networks (CNNs) are generally a type of deep learning model particularly effective for analyzing visual data, meaning images [35–38]. Below is an equation of a normal CNN model.

$$(f * g)(x, y) = \sum_{i=-k}^k \sum_{j=-k}^k f(x + i, y + j) \cdot g(i, j) \tag{11}$$

In this equation, f is the input image or feature map, representing the input image or the output from the previous layer in the CNN. Each value in this matrix corresponds to a pixel’s intensity in the case of an image or a feature in a feature map. g is the filter or kernel, representing a smaller matrix that slides over the input feature map. Each value in the filter corresponds to a weight that is applied to the input feature map. (x,y) is the position in the output feature map, k defines the size of the filter.

In the case of models with no images, the conventional CNN equation won’t work. In that case, a non-image equation, which uses 1-dimensional data (word/number-based datasets), would be utilized and is expressed as follows:

$$(f * g)(t) = \sum_{i=-k}^k f(t + i) \cdot g(i) \tag{12}$$

In this equation, f is the input sequence or time series data, representing a 1D array or a higher-dimensional tensor representing the input data. For instance, in time series analysis, each value could represent data at a specific time point. g is the filter or kernel,

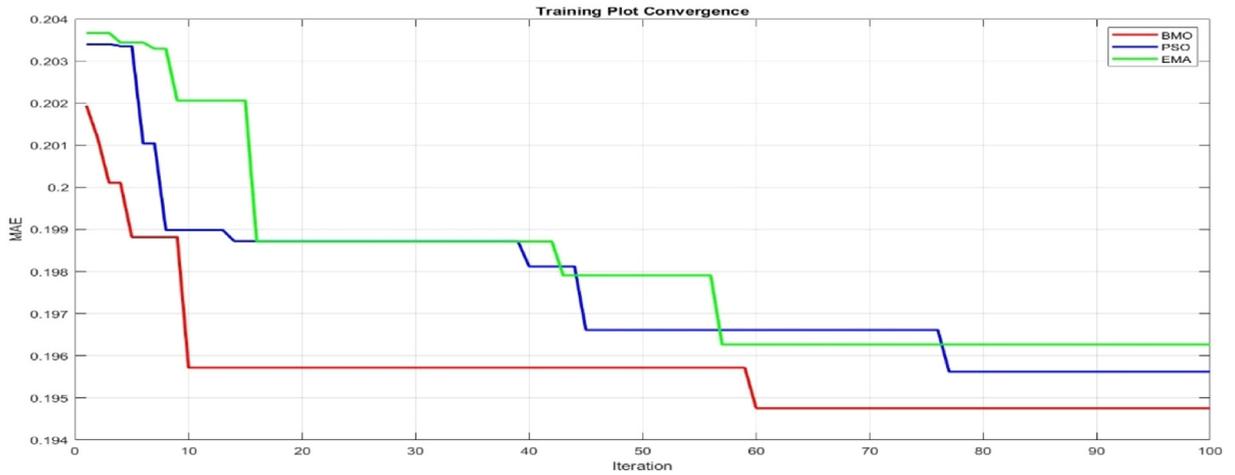


Fig. 4. Convergence Plot.

designed to detect specific patterns or features in the input data. t is the position in the output sequence whereas k defines the size of the filter.

$$z_t = f\left(\sum_{i=0}^{k-1} w_i \cdot x_{t+i} + b\right) \quad (13)$$

With regards to this paper, the deep learning model used is a hybrid of CNN, a Convolutional Neural Network combined with a Long Short-Term Memory network (CNN-LSTM). It functions by extracting features (CNN component) then sequence modeling it (LSTM component). It integrates the spatial feature extraction capabilities of CNNs, as shown in Eq. (13), where x_{t+i} represents the input data at certain points ($t + i$), w_i and b represents the weights and biases, k is the kernel size, f represents activation functions and z_t is the output, the feature extracted at certain points (t), with the temporal sequence modeling strengths of LSTMs. This combination is particularly effective for non-image sequential data where capturing both local patterns and long-term dependencies is crucial. After extracting the features, it would flow into the forget gate where it would determine which information would be discarded by understanding the amount of memory needed to be filtered which would then be processed to the input gate where it would select new information to be added. Then it would be released by the output gate where output based on the information provided by the input gate has been processed [39–41]

Statistical analysis

The performance metrics, mean absolute error, mean squared error, root mean squared error, standard deviation, coefficient of determination, and max error are utilized in this study for risk assessment while accuracy and precision for diagnosis assessment [42]. The models run under each partition have been compared to select which partition would be the best suited for this paper. The Wilcoxon signed-rank test was incorporated to statistically evaluate model performance differences, along with boxplots to illustrate the stability of the models across multiple test runs [43].

Method validation

The study was conducted using MATLAB on a laptop with an Intel Core i7 processor, Intel Iris Xe graphics, and 16GB RAM. Matlab was employed as the primary tool for processing the dataset and constructing the models.

Hyperparameters

The efficiency and convergence of each optimization of hyperparameters in each model are shown in Fig. 4. It shows the efficiency and convergence of the 60:40 partition as it was the best splitting ratio after trials had been completed. The results of this data partition can be inferred in Table 5. It shows how well each model's optimizer selects their hyperparameter for their model. It is to be noted that each model has been trained a minimum of 10 times. Table 4 shows the best parameters selected via each optimizer while Table 5 shows the statistical evaluation of the models after the hyperparameters are included in.

The convergence plot in Fig. 4 illustrates the optimization behavior of BMO, PSO, and EMA in minimizing the Mean Absolute Error (MAE) across 100 iterations. While BMO achieves the fastest and lowest training convergence, it is important to contextualize this with overall model performance. Despite EMA showing a slower and less optimal convergence curve—stabilizing at a slightly higher MAE compared to BMO during training—it ultimately outperforms both BMO and PSO in predictive evaluation metrics, as inferred in Table 5.

Table 4
Hyperparameters (Optimized).

Parameters	EMA	Fitness (MAE)	BMO	Fitness (MAE)	PSO	Fitness (MAE)
Number of filters	22.247	0.196	64	0.194	64	0.195
Filter size	3.749		4		4	
LSTM units	105.53		32		126	
Dropout rate	0.1		0.211		0.292	

Table 5
Optimized CNN-LSTM algorithms.

	CNN-LSTM-BMO	CNN-LSTM-PSO	CNN-LSTM-EMA	CNN-LSTM-EMA
At 60:40				5-fold Cross-validation
R ² (Risk Prediction)	0.93	0.95	0.98	0.99
MAE (Risk Prediction)	0.038	0.028	0.018	0.01
MSE (Risk Prediction)	0.002	0.0014	0.0006	0.0002
RMSE (Risk Prediction)	0.045	0.037	0.024	0.015
Max Error (Risk Prediction)	0.27	0.21	0.2	0.15
Std. Deviation (Risk Prediction)	0.032	0.034	0.023	0.015
Accuracy (Binary Prediction)	0.7	0.7	0.70	0.72
Precision (Binary Prediction)	0.74	0.7	0.8	0.8
Computational time	3 h and 25 min	3 h	4 hours	22 h

Statistical analysis results

The results presented in [Table 5](#) showcase a comparative evaluation of three optimization techniques—Barnacle Mating Optimization (BMO), Particle Swarm Optimization (PSO), and the Evolutionary Mating Algorithm (EMA)—based on their Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Coefficient of Determination (R²), Standard Deviation (STD) for the risk variable, accuracy and precision for the analysis variable and computational time for both variables at the best data partition ratio.

The CNN-LSTM-EMA model demonstrates superior performance in risk prediction across all evaluation metrics. With an R² of 0.98, it explains nearly all the variance in the data, outperforming CNN-LSTM-PSO (0.95) and CNN-LSTM-BMO (0.93). Its lower MAE (0.018), MSE (0.0006), and RMSE (0.024) indicate a more precise and consistent prediction capability, with minimal average and squared errors. Additionally, the lowest standard deviation (0.023) and maximum error (0.2) highlight the EMA-optimized model's robustness and reduced susceptibility to outlier-induced deviations. In contrast, the BMO-optimized model lags with the highest error values and variability, suggesting less reliable performance. While PSO offers moderate improvements over BMO, it still falls short of EMA's predictive accuracy and stability. For binary diagnosis, all three models achieve comparable accuracy (0.7). However, the CNN-LSTM-EMA model attains the highest precision (0.8), indicating better performance in correctly identifying positives and reducing false positives. BMO (precision = 0.74) and PSO (0.7) follow, with diminishing precision.

When extended to a 5-fold cross-validation framework, the CNN-LSTM-EMA model achieves even stronger generalization, recording an R² of 0.99, MAE of 0.01, MSE of 0.0002, RMSE of 0.015, StD of 0.015, and Max error of 0.15. For binary diagnosis, the model attains a marginally improved accuracy of 0.72, while maintaining the highest precision of 0.80 among all tested variants. These results confirm EMA's superior capability in hyperparameter tuning, particularly in avoiding local minima and enabling effective exploration of the optimization landscape.

While the 5-fold cross-validated CNN-LSTM-EMA model provides the highest performance metrics, it incurs a computational time of approximately 22 h, which may limit its feasibility for real-time or large-scale clinical deployment. In contrast, the 60/40 train-test split CNN-LSTM-EMA model completes in a significantly shorter computational time of approximately 4 h, while still achieving highly competitive performance. This renders it a more pragmatic option for applications where computational efficiency is critical.

This indicates that faster convergence does not necessarily correlate with superior generalization. EMA's relatively higher MAE during training is compensated by its enhanced capability to escape local minima and achieve better global optimization of hyperparameters. This results in the CNN-LSTM-EMA model achieving the best R² scores, lowest RMSE, and superior precision (0.8) in actual testing scenarios.

Prediction plots

The Evolutionary Mating Algorithm (EMA) has demonstrated significant advantages in optimizing CNN-LSTM models, particularly in terms of achieving better distribution and precision in risk predictions. When compared to other optimization methods, EMA ensures a closer alignment between the predicted and actual risk outputs, as shown in [Fig. 5](#). The bar plot (a snapshot of the complete plot) provides a comparative visualization of actual versus predicted cardiovascular risk values for a subset of test samples, highlighting the performance of CNN-LSTM models optimized by EMA, BMO, and PSO. It is to be noted that the bar plot did not include the 5-fold cross-validated EMA. Across the selected indices, the CNN-LSTM-EMA model (orange) consistently demonstrates the closest alignment

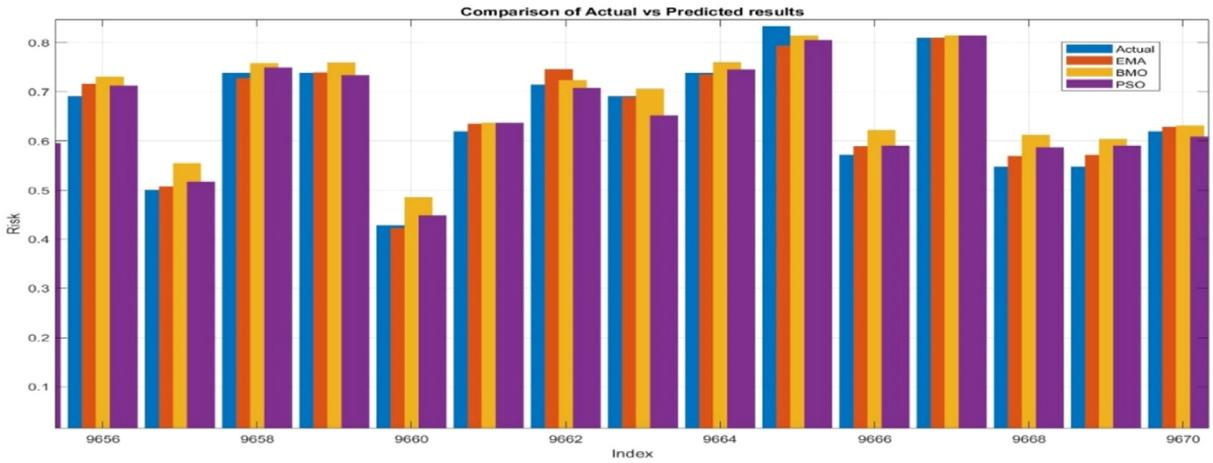


Fig. 5. Risk assessment Actual vs Predicted.

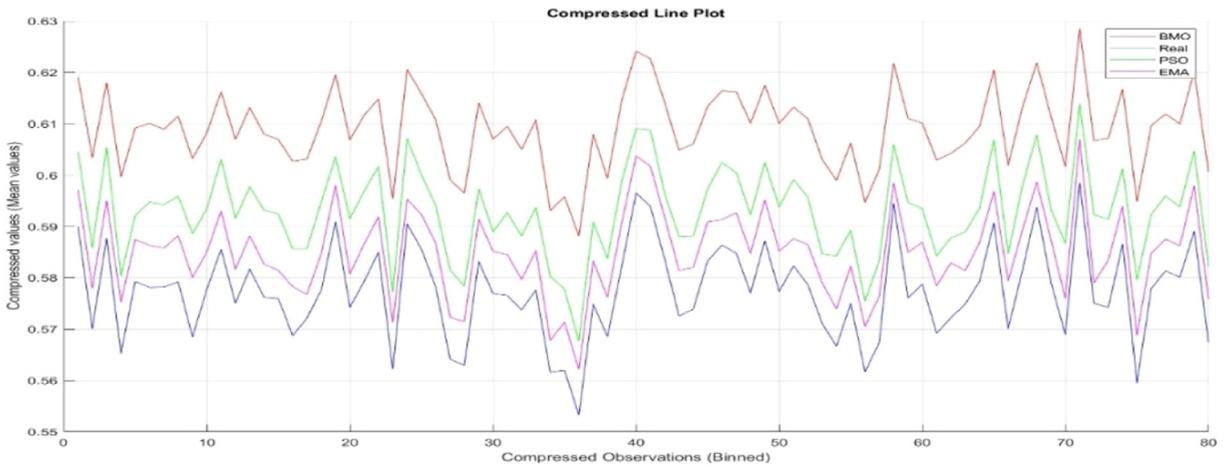


Fig. 6. Risk assessment Actual vs Predicted (Line Plot).

with the actual risk values (blue), reinforcing its superior predictive accuracy and reduced deviation. In contrast, the CNN-LSTM-BMO model (yellow) shows more pronounced over- or under-estimations, suggesting a less reliable generalization. The CNN-LSTM-PSO model (purple) performs moderately, offering improved accuracy over BMO but still falling short of EMA’s precision.

These visual trends corroborate earlier statistical findings, where EMA achieved the lowest MAE (0.018), MSE (0.0006), and RMSE (0.024), along with the highest R^2 (0.98). The bar plot thus visually substantiates the quantitative evaluation, validating EMA as the most reliable optimization strategy among those evaluated.

The provided line plot in Fig. 5 is a compressed version of the earlier bar plot in Fig. 5, offering a continuous visualization of the prediction performance of all the models compared to actual values with the actual values represented by blue, CNN-LSTM-BMO represented by red, CNN-LSTM-PSO represented by green and CNN-LSTM-EMA represented by magenta. This plot allows for a clearer comparison of how well each predictive model tracks the actual risk values over a broader dataset. The compressed format of the plot provides an overall trend but loses some granularity compared to the bar plot in Fig. 5. While the bar plot facilitated direct comparisons at specific observations, the line plot in Fig. 6 effectively illustrates fluctuations over a broader range.

Statistical significance testing and feature importance

Fig. 7 depicts the feature importance of all the variables utilized in the optimized CNN-LSTM model and how it impacts the results acquired. Achieved through Shapley Additive explanations (SHAP) analysis. Fig. 7 above presents a SHAP analysis for feature importance, illustrating the contribution of different features to the model’s predictive capability. From visualization, the feature Age exhibits the highest SHAP value, signifying its dominant role in the model’s decision-making process. This suggests that this feature has the most significant impact on risk prediction.

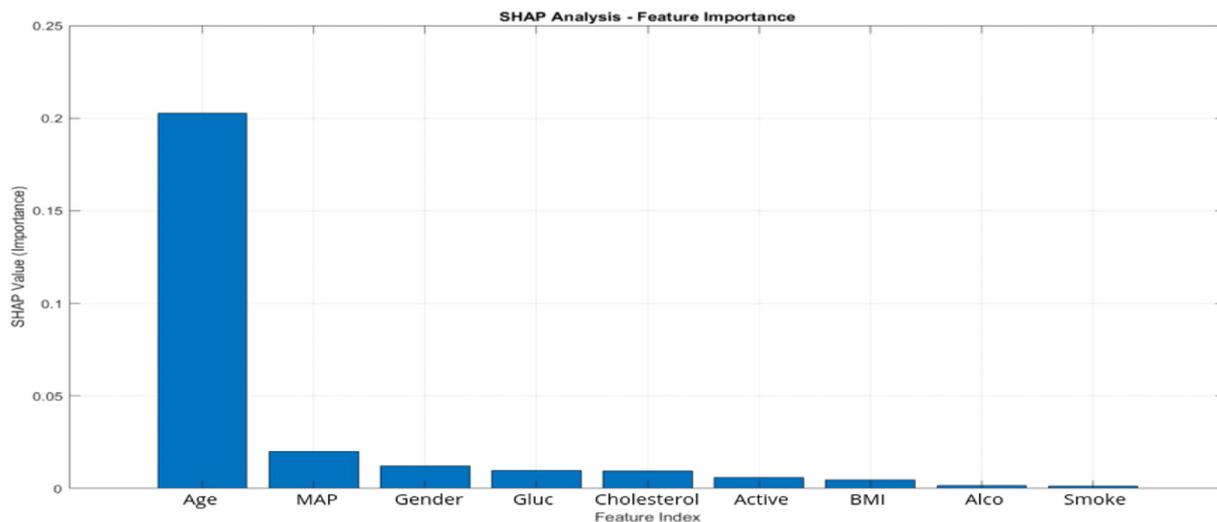


Fig. 7. SHAP Analysis- Feature importance.

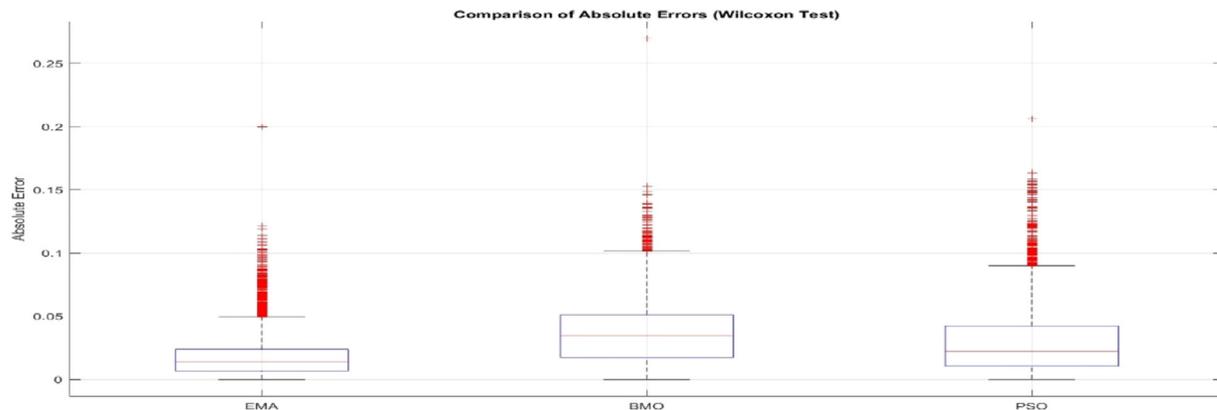


Fig. 8. Boxplot (significance test).

Fig. 8 presents a boxplot comparison of absolute errors for three models, evaluated using the Wilcoxon signed-rank test. The x-axis represents the three models, while the y-axis denotes the absolute error values. The boxplots display the median (red line), interquartile range (IQR) (blue box), and outliers (red points), offering insight into the distribution and variability of absolute errors across different models.

From the visualization, Model 1 (EMA-optimized) demonstrates the most favorable error distribution, characterized by a narrow interquartile range, low median absolute error, and minimal variability—indicating stable and consistent predictive performance. Model 3 (PSO-optimized) follows though it presents slightly higher dispersion and a greater number of outliers. In contrast, Model 2 (BMO-optimized) exhibits a wider spread and higher median error, reflecting reduced robustness. The presence of outliers across all models, most notably in Model 3—highlights occasional deviations, yet EMA's tighter error bounds reinforce its superior reliability in risk prediction.

The Wilcoxon signed-rank test, a non-parametric statistical test used to compare paired samples, was employed to determine whether significant differences exist in the absolute error distributions between the models. The obtained p-values for all three comparisons were <0.05 , and the test statistic $h = 1$, indicating that the differences between the models are statistically significant. This suggests that at least one model significantly outperforms the others in terms of error minimization. Given these findings, Models 1 and 3 demonstrate superior predictive accuracy compared to Model 2.

Limitations

The findings in this paper underscore EMA's utility in optimizing deep learning models for multifaceted medical prediction tasks, enabling both precise risk assessment and reliable disease diagnostic classification. However, computational demands remain a notable limitation, as does the model's slightly higher performance bias toward the regression component.

While the 5-fold cross-validated CNN-LSTM-EMA configuration yields the highest performance metrics, it incurs significantly longer training times compared to the more efficient 60/40 split variant. This challenge could be addressed in future studies by incorporating further algorithmic refinements or leveraging high-performance computing environments to reduce runtime without compromising accuracy. Future work should address these limitations through architectural enhancements and advanced multi-objective optimization techniques. To address the performance bias, future work could mitigate it through architectural enhancements, such as the integration of multi-task learning frameworks and customized loss functions (e.g., weighted or multi-objective loss functions) to facilitate equitable learning across heterogeneous output types and enhance the model's capacity for balanced optimization. To ensure clinical applicability, future research should focus on validating the model across diverse populations and institution-specific and real-time datasets such as real-world electronic health records (EHRs) systems. To further strengthen predictive performance and real-world relevance, future research could incorporate multi-modal data, including ECG signals, imaging, and genetic biomarkers, to further improve predictive robustness and potentially improve diagnostic precision, advancing the development of AI-driven tools for early CVD detection and personalized patient care.

Limitations in this paper include the following:

1. Its computational cost, which could be mitigated by optimizing the algorithm further or utilizing high-performance computing resources.
2. Due to its nature of predicting 2 outputs of different types (continuous/regression for risk and binary for diagnosis), the model mostly favors the regression type, thus not performing as well for the second type of output. This could be mitigated by further optimizing the algorithm to better process dual outputs.
3. While the proposed model demonstrates strong predictive performance, its real-world deployment requires further validation across different populations to ensure fairness and reliability.

Ethics statements

The data used in this study is a publicly available medical dataset obtained from Kaggle. All data were anonymized prior to release, and no personal identifiable information was included. The use of this data complies with the terms of use provided by the original data providers.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work was supported by the Ministry of Higher Education Malaysia (MOHE) under Fundamental Research Grant Scheme (FRGS/1/2022/ICT04/UMP/02/1) and Universiti Malaysia Pahang (#RDU220105).

References

- [1] WHO, Noncommunicable diseases. World Health Organization (2024). <https://www.who.int/health-topics/noncommunicable-diseases>.
- [2] National Center for, Chronic Disease Prevention and Health Promotion, "Heart disease and stroke (2024). <https://www.cdc.gov/chronicdisease/resources/publications/factsheets/heart-disease-stroke.htm>.
- [3] R. Sigit, R. Rokhana, Setiawardhana, T. Hidayat, Anwar, J.J. Jaenputra, Implementation of portable ultrasound for heart disease detection using cloud computing-based machine learning, *EMITTER Int. J. Eng. Technol.* 12 (2) (2024) 196–212 Dec., doi:10.24003/EMITTER.V12I2.904.
- [4] D. Peng, J. Bashir, J. Abel, J. Ye, Long-term outcomes of surgical aortic valve replacement in patients receiving chronic dialysis, *Canad. J. Surg.* 66 (4) (2023) E358 Jul., doi:10.1503/CJS.019121.
- [5] M. Fatima, M. Pasha, Survey of machine learning algorithms for disease diagnostic, *J. Intellig. Learn. Syst. Applic.* 09 (01) (2017) 1–16, doi:10.4236/jilsa.2017.91001.
- [6] P. Natarajan, J. Frenzel, D. Smaltz, *Demystifying Big Data and Machine Learning For Healthcare*, 1st edition, CRC Press, 2021.
- [7] M. Ordikhani, M.S. Abadeh, C. Prugger, R. Hassannejad, N. Mohammadifard, N. Sarrafzadegan, An evolutionary machine learning algorithm for cardiovascular disease risk prediction, *PLoS. One* 17 (7) (2022) e0271723 Jul., doi:10.1371/JOURNAL.PONE.0271723.
- [8] E. Tjoa, C. Guan, A survey on explainable artificial intelligence (XAI): toward medical XAI, *IEEe Trans. Neural Netw. Learn. Syst.* 32 (11) (2021) 4793–4813 Nov., doi:10.1109/TNNLS.2020.3027314.
- [9] E. Özbay, An active deep learning method for diabetic retinopathy detection in segmented fundus images using artificial bee colony algorithm, *Artif. Intell. Rev.* 56 (4) (2023) 3291–3318 Apr., doi:10.1007/S10462-022-10231-3.
- [10] X. Wang, Z. Shi, X. Ji, B. Hu, S. Chen, L. Zhang, Nationwide survey of the status of artificial intelligence-based intracranial aneurysm detection systems, *Intell. Med.* (2024) Dec., doi:10.1016/J.IMED.2024.11.001.
- [11] S.A. Ali Shah, I. Uddin, F. Aziz, S. Ahmad, M.A. Al-Khasawneh, M. Sharaf, An enhanced deep neural network for predicting workplace absenteeism, *Complexity*. 2020 (2020) 1–12 Feb., doi:10.1155/2020/5843932.
- [12] P.M. Shanthappa, M. Bayari, G.B. Abhilash, K.V. Gokul, P.J. Ashish, Parkinson's disease detection using inceptionV3: a deep learning approach, *MethodsX*. 14 (2025) 103333 Jun., doi:10.1016/J.MEX.2025.103333.
- [13] O. Deperlioglu, U. Kose, D. Gupta, A. Khanna, A.K. Sangaiah, Diagnosis of heart diseases by a secure Internet of Health Things system based on Autoencoder Deep Neural Network, *Comput. Commun.* 162 (2020) 31–50, doi:10.1016/j.comcom.2020.08.011.

- [14] B.P. Manoj Kumar, S.R. Perumal, N.R. K. Type 2: diabetes mellitus prediction using Deep Neural Networks classifier, *Int. J. Cognit. Comput. Eng.* 1 (2020) 55–61, doi:10.1016/j.ijcce.2020.10.002.
- [15] P. Nawghare, J. Prasad, Hybrid CNN and random forest model with late fusion for detection of autism spectrum disorder in toddlers, *MethodsX*. 14 (2025) 103278 Jun., doi:10.1016/J.MEX.2025.103278.
- [16] S. Edwin Raja, J. Sutha, P. Elamparithi, K.Jaya Deepthi, S.D. Lalitha, Liver tumor prediction using attention-guided convolutional neural networks and genomic feature analysis, *MethodsX*. 14 (2025) 103276 Jun., doi:10.1016/J.MEX.2025.103276.
- [17] A. Kumar, V.K. Patel, Classification and identification of disease in potato leaf using hierarchical based deep learning convolutional neural network, *Multimedia Tools Appl.* 82 (20) (2023) 31101–31127 Aug., doi:10.1007/s11042-023-14663-z.
- [18] F.A. Mohammed, K.K. Tune, B.G. Assefa, M. Jett, S. Muhie, Medical image classifications using convolutional neural networks: a survey of current methods and statistical modeling of the literature, *Mach. Learn Knowl. Extr.* 6 (1) (2024) 699–735 Mar., doi:10.3390/MAKE6010033/S1.
- [19] M. Lin, et al., A combined system with convolutional neural networks and transformers for automated quantification of left ventricular ejection fraction from 2D echocardiographic images, *Intell. Med.* (2024) Dec., doi:10.1016/J.IMED.2024.10.001.
- [20] E.S.M. El-kenawy, N. Khodadadi, S. Mirjalili, A.A. Abdelhamid, M.M. Eid, A. Ibrahim, Greylag Goose Optimization: nature-inspired optimization algorithm, *Expert. Syst. Appl.* 238 (2024) 122147 Mar., doi:10.1016/J.ESWA.2023.122147.
- [21] C. Vidal, et al., Robust xEV battery State-of-charge estimator design using a feedforward deep neural network, *SAE Int. J. Adv. Curr. Pract. Mobil.* 2 (5) (2020) 2872–2880 Apr., doi:10.4271/2020-01-1181.
- [22] C. Vidal, P. Malysz, M. Naguib, A. Emadi, P.J. Kollmeyer, Estimating battery state of charge using recurrent and non-recurrent neural networks, *J. Energy Storage* 47 (2022) 103660 Mar., doi:10.1016/J.EST.2021.103660.
- [23] M.H. Sulaiman, Z. Mustafa, M.M. Saari, H. Daniyal, S. Mirjalili, Evolutionary mating algorithm, *Neural Comput. Appl.* 35 (1) (2023) 487–516, doi:10.1007/s00521-022-07761-w.
- [24] A. Suresh, .R. Kumar, .R. Varatharajan, B.A. Suresh, R. Kumar, R. Varatharajan, Health care data analysis using evolutionary algorithm, *J. Supercomput.* 76 (2020) 4262–4271, doi:10.1007/s11227-018-2302-0.
- [25] Y. Clapper, J. Berkhou, R. Bekker, D. Moeke, A model-based evolutionary algorithm for home health care scheduling, *Comput. Oper. Res.* 150 (2023) 106081, doi:10.1016/j.cor.2022.106081.
- [26] M. Mahmoud, A review on waste management techniques for sustainable energy production, *Metaheurist. Optimiz. Rev.* 3 (2) (2025) 47–58, doi:10.54216/MOR.030205.
- [27] S. Ulianova, Cardiovascular disease dataset (2024). <https://www.kaggle.com/datasets/sulianova/cardiovascular-disease-dataset>.
- [28] H. Komasi, S. Zolfani, J. Antucheviciene, Emphasizing access to health and treatment services in order to identify the key factors influencing an age-friendly city, *Int. J. Strateg. Property Manag.* 27 (2023) 176–187 Jan., doi:10.3846/ijspm.2023.19865.
- [29] H. Islam, M.S. Iqbal, M.M. Hossain, Blood pressure abnormality detection and interpretation utilizing explainable artificial intelligence, *Intell. Med.* (2024) Nov., doi:10.1016/J.IMED.2024.09.005.
- [30] National Institute on Aging, Heart health and aging, 2025. <http://www.nia.nih.gov/health/heart-health/heart-health-and-aging>.
- [31] K. Drożdż, et al., Risk factors for cardiovascular disease in patients with metabolic-associated fatty liver disease: a machine learning approach, *Cardiovasc. Diabetol.* 21 (1) (2022) 240 Nov., doi:10.1186/s12933-022-01672-9.
- [32] F. García-Ávila, T. Encalada-Alvear, L.M. Alvarez-Altamirano, A. Avilés-Añazco, L. Valdiviezo-Gonzalez, E. Alfaro-Paredes, Human health risk assessment of aluminium in water treated with aluminium-based coagulants in a rural area, *Results. Eng.* (2025) 104200 Jan., doi:10.1016/J.RINENG.2025.104200.
- [33] R. Waigi, S. Choudhary, P. Fulzele, G. Mishra, Predicting the risk of heart disease using advanced machine learning approach, *Europ. J. Molec. Clin. Med.* 7 (2020) 1638–1645 Jan..
- [34] A.-P. Kengne, et al., Blood pressure variables and cardiovascular risk, *Hypertension* 54 (2) (2009) 399–404 Aug., doi:10.1161/HYPERTENSIONAHA.109.133041.
- [35] T. Zhou, Q.R. Cheng, H.L. Lu, Q. Li, X.X. Zhang, S. Qiu, Deep learning methods for medical image fusion: a review, *Comput. Biol. Med.* 160 (2023) 106959 Jun., doi:10.1016/J.COMPBIOMED.2023.106959.
- [36] A.M. Ayalew, A.O. Salau, Y. Tamyalew, B.T. Abeje, N. Woreta, X-ray image-based COVID-19 detection using deep learning, *Multimed. Tools. Appl.* 82 (28) (2023) 44507–44525 Nov., doi:10.1007/s11042-023-15389-8/FIGURES/13.
- [37] Y. Li, J. Zhao, Z. Lv, J. Li, Medical image fusion method by deep learning, *Int. J. Cognit. Comput. Eng.* 2 (2021) 21–29 Jun., doi:10.1016/J.IJCCCE.2020.12.004.
- [38] T.Ayoub Shaikh, T. Rasool, F.Rasheed Lone, Towards leveraging the role of machine learning and artificial intelligence in precision agriculture and smart farming, *Comput. Electron. Agric.* 198 (2022) 107119 Jul., doi:10.1016/j.compag.2022.107119.
- [39] F. Elmaz, R. Eyckerman, W. Casteels, S. Latré, P. Hellinckx, CNN-LSTM architecture for predictive indoor temperature modeling, *Build. Environ.* 206 (2021) 108327 Dec., doi:10.1016/J.BUILDENV.2021.108327.
- [40] M.F. Hosen, et al., An LSTM network-based model with attention techniques for predicting linear T-cell epitopes of the hepatitis C virus, *Results. Eng.* 24 (2024) 103476 Dec., doi:10.1016/J.RINENG.2024.103476.
- [41] S. Padhy, S. Dash, N. Kumar, S.P. Singh, G. Kumar, P. Moral, Temporal integration of ResNet features with LSTM for enhanced skin lesion classification, *Results. Eng.* (2025) 104201 Jan., doi:10.1016/J.RINENG.2025.104201.
- [42] K. Khaled, M.K. Singla, Predictive analysis of groundwater resources using random forest regression, *J. Artific. Intellig. Metaheurist.* 09 (01) (2025) 11–19, doi:10.54216/JAIM.090102.
- [43] E.S.M. El-kenawy, et al., Sunshine duration measurements and predictions in Saharan Algeria region: an improved ensemble learning approach, *Theor. Appl. Climatol.* 147 (3–4) (2022) 1015–1031 Feb., doi:10.1007/S00704-021-03843-2/METRICS.