Journal of Electronics, Electromedical Engineering, and Medical Informatics
Homepage: jeeemi.org; Vol. 7, No. 3, July 2025, pp: 713-728 **e-ISSN: 2656-8632**

RESEARCH ARTICLE

OPEN ACCESS

Performance Comparison of Extreme Learning Machine (ELM) and Hierarchical Extreme Learning Machine (H-ELM) Methods for Heart Failure Classification on Clinical Health Datasets

Ichwan Dwi Nugraha[1], Triando Hamonangan Saragih[1], Irwan Budiman[1], Dwi Kartini[1] ,Fatma Indriani[1], and Wahyu Caesarendra[2]

[1] Computer Science Department, Lambung Mangkurat University, Banjarbaru, South Kalimantan, Indonesia
[2] Department of Mechanical and Mechatronics Engineering, Faculty of Engineering and Science, Curtin University Malaysia, Sarawak, Malaysia

Corresponding author: Triando Hamonangan Saragih. (e-mail: triando.saragih@ulm.ac.id), **Author(s) Email**: Ichwan Dwi Nugraha (e-mail: ichwandwinugraha@gmail.com), Irwan Budiman (e-mail: irwan.budiman@ulm.ac.id), Dwi Kartini (e-mail: dwikartini@ulm.ac.id), Fatma Indriani (e-mail: f.indriani@ulm.ac.id), Wahyu Caesarendra (wahyucaesarendra@gmail.com)

Abstract **Heart failure is one of the leading causes of death worldwide and requires accurate and timely diagnosis to improve patient outcomes. However, early detection remains a significant challenge due to the complexity of clinical data, high dimensionality of features, and variability in patient conditions. Traditional clinical methods often fall short in identifying subtle patterns that indicate early stages of heart failure, motivating the need for intelligent computational techniques to support diagnostic decisions. This study aims to enhance predictive modeling for heart failure classification by comparing two supervised machine learning approaches: Extreme Learning Machine (ELM) and Hierarchical Extreme Learning Machine (HELM). The main contribution of this research is the empirical evaluation of HELM's performance improvements over conventional ELM using 10-fold cross-validation on a publicly available clinical dataset. Unlike traditional neural networks, ELM offers fast training by randomly assigning weights and analytically computing output connections, while HELM extends this with a multi-layer structure that allows for more complex feature representation and improved generalization. Both models were assessed based on classification accuracy and Area Under the Curve (AUC), two critical metrics in medical classification tasks. The ELM model achieved an accuracy of 73.95% ± 8.07 and an AUC of 0.7614 ± 0.093, whereas the HELM model obtained a comparable accuracy of 73.55% ± 7.85 but with a higher AUC of 0.7776 ± 0.085. In several validation folds, HELM outperformed ELM, notably reaching 90% accuracy and 0.9250 AUC in specific cases. In conclusion, HELM demonstrates improved robustness and discriminatory capability in identifying heart failure cases. These findings suggest that HELM is a promising candidate for implementation in clinical decision support systems. Future research may incorporate feature selection, hyperparameter optimization, and evaluation across multi-center datasets to improve generalizability and real-world applicability.**

Keywords **Extreme Learning Machine; Hierarchical Extreme Learning Machine; Heart Failure.**

I. Introduction

Heart failure continues to be a major global health concern, contributing to more than 17 million deaths each year due to cardiovascular conditions, as reported by the World Health Organization in 2021 [1], making it one of the most critical diseases of the 21st century [2]. Timely detection is essential for reducing complications and enhancing patient prognosis [3]. However, traditional clinical diagnostic methods are often limited by the complexity and high dimensionality of medical data, which can hinder accurate and timely predictions [4].

In response to these challenges, machine learning (ML) and deep learning approaches have shown

Manuscript received March 10, 2025; Revised May 17, 2025; Accepted May 20, 2025; date of publication June 1, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.904

promising results in analyzing large-scale and complex clinical datasets that can help clinical researchers to target high-risk patients to modify treatment interventions based on accurate risk predictions [5], [6]. Deep learning methods can discover hidden patterns in patient data that are difficult to detect using conventional techniques, leading to more accurate early-stage diagnosis and targeted intervention [7]. One algorithm that has gained attention in this domain is the Extreme Learning Machine (ELM), which is known for its fast training and strong generalization performance in classification problems [8], [9], [10]. While ELM offers efficiency and simplicity, its performance may be limited in handling high-dimensional or nonlinear data structures. To address this, a more advanced model, the Hierarchical Extreme Learning Machine (HELM), has been proposed. HELM employs a multi-layer ELM-based architecture, allowing hierarchical feature representation that can capture more complex patterns [11], [12]. This approach has demonstrated improved accuracy in several domains, including medical classification tasks involving pathological voices and clinical diagnosis.

Despite the growing interest in both ELM and HELM, comparative studies assessing their predictive performance on clinical heart failure datasets are still limited. Prior works have largely focused on optimization techniques for ELM (e.g., using Particle Swarm Optimization), or on standalone evaluations of HELM for other health applications. Therefore, a clear research gap exists in evaluating and benchmarking these two models side by side on heart failure prediction tasks.

This study aims to fill the identified research gap by conducting a comprehensive comparative analysis between Extreme Learning Machine (ELM) and Hierarchical Extreme Learning Machine (HELM) using the publicly available Heart Failure Clinical Records Dataset. To ensure robust evaluation, we apply a 10-fold cross-validation framework and assess the predictive performance of both models using classification accuracy and Area Under the Curve (AUC) metrics. The contributions of this study include: (1) the implementation and benchmarking of ELM and HELM for heart failure classification, (2) the use of a cross-validation protocol to ensure generalizability of the results, (3) an empirical comparison of key evaluation metrics including accuracy and AUC, and (4) a critical discussion on the implications and potential of hierarchical learning architectures in the context of clinical prediction tasks.

II. Methods

This study follows a structured methodology comprising data collection, preprocessing, model implementation, and evaluation. The overall research workflow is illustrated in Fig. 1, starting from dataset preparation through model training and performance assessment.

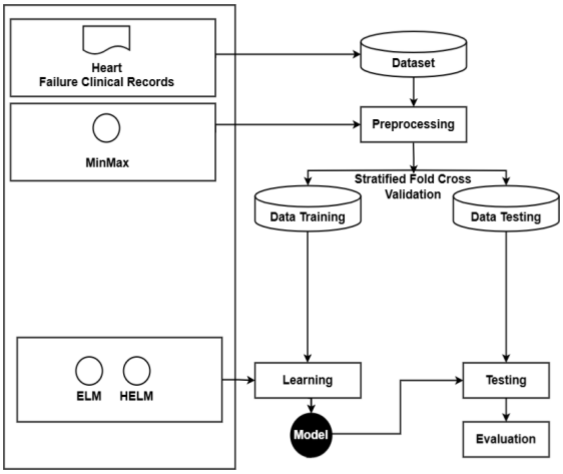

Fig. 1. This figure illustrates the research methodology workflow for heart failure diagnosis, detailing the sequential steps from clinical data collection and preprocessing through model learning and rigorous evaluation

A. Dataset

In Table 1 the dataset used in this study is the Heart Failure Clinical Records Dataset, which is publicly available on Kaggle (https://www.kaggle.com/code/karnikakapoor/heartfailure-prediction-ann). This dataset contains 299 patient records with 13 clinical features and one binary target variable (DEATH_EVENT) indicating whether the patient died during the follow-up period. The features include demographic, clinical, and laboratory variables such as age, sex, ejection fraction, serum creatinine, smoking status, and more. All records are derived from real patients who had left ventricular systolic dysfunction, providing a reliable basis for heart failure prediction research.

Table 1. Summary of heart failure clinical dataset consisting of 299 patients and 13 attributes, including the DEATH_EVENT outcome.

Age	Anaemia	Creatinine_phosphokinase	Diabetes	…	Death_event
75	0	582	0	…	1
55	0	7861	0	…	1
65	0	146	0	…	1
50	1	111	0	…	1
65	1	160	1	…	

Manuscript received March 10, 2025; Revised May 17, 2025; Accepted May 20, 2025; date of publication June 1, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.904

90	1	47	0	…	1
75	1	246	0	…	1
60	1	315	1	…	1
65	0	157	0	…	1
80	1	123	0	…	1
75	1	81	0	…	1
62	0	231	0	…	1
…	…	…	…	…	…
50	0	196	0	45	0

B. Data Preprocessing

Before model training, the dataset was preprocessed through several stages: feature-label separation, handling of missing values (if any), and normalization [13]. The process of normalization is essential in order to mitigate the substantial impact of variables that exhibit significantly disparate value ranges [14]. At present, with the rapid acceleration and expansion of data, the presence of missing values is a prevalent phenomenon in quantitative research [15]. Normalization is necessary to ensure all features operate within the same scale. In this study, Min-Max normalization was applied as expressed in Eq. (1) [16], [17], [18].

$$x^1 = \frac{x_i - \min(x)}{\max(x) - \min(x)} \quad (1)$$

This transformation improves convergence and prevents any one feature from dominating the model during training. Where X1 signifies a unique value designated for the purpose of normalization, x1 represents the resultant value subsequent to the normalization process, minx(x) denotes the minimum value pertinent to a specific attribute, whereas max(x) indicates the maximum value associated with an attribute. The range is confined within the interval [0,1], and the length of this interval is quantified as 1[19].

Prior to model training, several preprocessing steps were applied to the dataset. First, the dataset was examined for missing values. No missing data was found, so no imputation was necessary. Next, categorical features such as sex, smoking, diabetes, and anaemia were already in binary format and thus did not require encoding. To bring all numerical features into the same scale, Min-Max normalization was used, transforming feature values to a [0, 1] range. This step is important to prevent features with larger ranges from dominating model training. No feature elimination or dimensionality reduction was conducted, as all features in the dataset were clinically relevant and consistently used in previous heart failure studies.

C. Cross-validation

To ensure the robustness and generalizability of the model, we adopted 10-fold cross-validation [20]. In this study, Stratified K-Fold Cross-Validation with 10 folds was used to ensure balanced class distribution across all folds. Each fold preserves the proportion of patients who experienced a death event and those who did not. The dataset was randomly shuffled before splitting, and the random state was set to 42 for reproducibility. This cross-validation approach helps ensure that the reported model performance is not biased due to data partitioning. The dataset was split into 10 equal parts. In each iteration, 9 folds were used for training and 1 fold for testing, rotating the test set until all folds had been used. This process reduces overfitting and yields an average performance score across all folds [21], as shown in Fig. 2 . A schematic representation of the ten-fold cross-validation methodology is delineated. The dataset was methodically divided into ten discrete segments, with nine segments employed as training data in an iterative framework, while one segment was allocated for evaluative purposes as test data. The mean value E of the outcomes derived from the ten segments is computed to approximate the model's accuracy and functions as a quantitative metric for the evaluation of the prevailing K-fold cross-validation framework. In this particular context, Ei denotes the cross-validation error associated with the ith segment [22]. The predominant constraint of K-Fold Cross Validation is ascribed to the disproportionate distribution of data, which engenders a potential risk of data loss, particularly accentuated when interacting with imbalanced datasets[23].

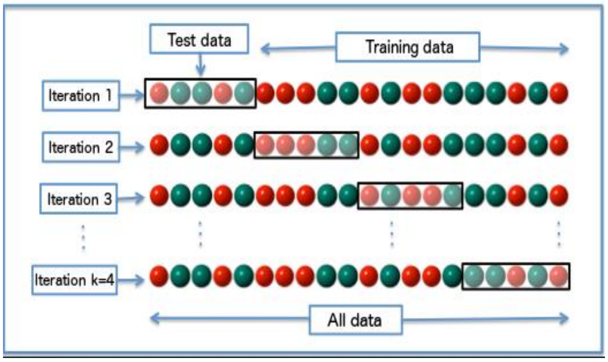

Fig. 2. Cross-validation stages: partition data, rotate roles, assess performance.

D. Model Implementation

In the present investigation, two distinct supervised learning paradigms were instantiated and subjected to comparative analysis: the Extreme Learning Machine (ELM) and the Hierarchical Extreme Learning Machine (HELM). The selection of these models is predicated upon their established efficacy in a plethora of pattern

Manuscript received March 10, 2025; Revised May 17, 2025; Accepted May 20, 2025; date of publication June 1, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.904

recognition applications, with a particular emphasis on biomedical data analysis [24], [25]. ELM is extensively acknowledged for its inherent simplicity and computational efficiency, particularly in the context of managing small to medium-sized datasets characterized by high-dimensional features [26]. Conversely, HELM is engineered to address certain shortcomings inherent in shallow learning methodologies by utilizing a hierarchical architecture that facilitates the extraction of more abstract and discriminative features across multiple layers.

The Extreme Learning Machine (ELM) constitutes a single-hidden-layer feedforward neural network (SLFN) wherein the input weights and hidden layer biases are assigned randomly and remain constant throughout the training process [27], [28].

This architecture is composed of an input layer, a hidden layer, and an output layer, thus eliminating the need for iterative training [29]. Only the output weights are acquired through analytical means employing a least-squares technique, which markedly diminishes the training duration when juxtaposed with traditional neural networks [8]. Such attributes render ELM particularly amenable to real-time applications necessitating rapid training, as evidenced in [30] Fig. 3. In relation to the overarching computational framework for single-layer Extreme Learning Machines (ELM), the training dataset is denoted by X: a collection of N labeled pairs (xi,yi)(xi,yi), where $xi\in R$xi∈R signifies the ith input vector and $yi\in R$yi∈R represents the corresponding anticipated "target" value. Consequently, to ascertain the output, the function $f(x)$f(x) pertinent to an ELM is articulated as delineated in Eq. (2) [31].

$$f(x) = \sum_{i=1}^{M} w_j \cdot a\left(r_j \cdot x + b_j\right) \tag{2}$$

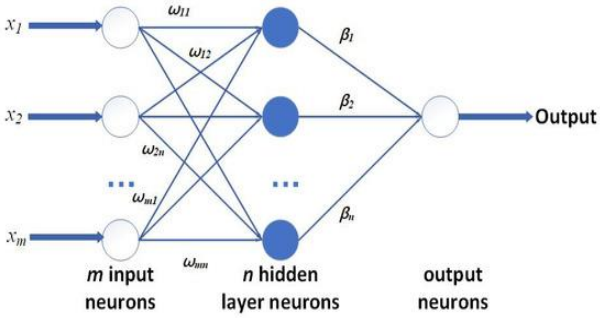

Fig. 3. Illustrates the structure of the Extreme Learning Machine (ELM), which is based on a single-hidden-layer feedforward neural network architecture

Within the ELM framework, the input weights (w) and biases (b) of the hidden layer are assigned randomly and remain unaltered throughout the training phase. Each hidden neuron applies a nonlinear activation function to a linear amalgamation of inputs (hi(x) = g(wix + b_1)), culminating in the formation of a hidden layer output matrix H. The output weights (8) that interconnect the hidden layer with the output layer are subsequently computed using the Moore-Penrose pseudoinverse, expressed as ß = H+T, where T denotes the target output. This methodological approach enables ELM to attain competitive classification efficacy while significantly curtailing training duration in comparison to conventional neural networks.

The input layer, characterized by M neurons, is interconnected with the hidden layer composed of H neurons via a specified set of weights, $rj\in R,j=1,\dots,H$rj∈R,j=1,…,H, while the jth hidden neuron incorporates a bias term designated as bjbj and a nonlinear activation function symbolized by $a(\cdot)$a(▪). A vector encompassing weighted connections, $w\in R$w∈R, computes the output neuron in conjunction with the hidden layer. The parameters rj,bjrj, bj as delineated in (1) are initialized randomly and remain unoptimized. The parameters rj,bjrj,bj as indicated in (1) are initialized randomly and remain unoptimized. Let H represent the activation matrix, wherein $hij\in H(i=1,\dots,N;j=1,\dots,H)$hij∈H(i=1,…,N;j=1,…,H) signifies the activation value of the jth hidden neuron corresponding to the ith input pattern, articulated as $hij=a(rj\cdot xi+bj)$hij=a(rj▪xi+bj).

A comprehensive elucidation of results is presented in Eq. (3) [31].

$$\begin{aligned} &H(w_1,\dots,w_n,b_1,\dots,b_n,x_1,\dots,x_n) \\ &= \begin{bmatrix} a(w_1x_1+b_1) & \cdots & a(w_nx_1+b_n) \\ \vdots & \ddots & \vdots \\ a(w_1x_N+b_1) & \cdots & a(w_nx_N+b_n) \end{bmatrix} \end{aligned} \tag{3}$$

In general, the training process of the ELM will reach completion upon the optimization of the following convex cost function as expressed in Eq. (4) [31].

$$\min_{(w,b)} |Hw - \mathcal{Y}|^2 \tag{4}$$

In general, Extreme Learning Machine (ELM) training can be done through the following steps: First, randomly assign parameters rj and bj to each hidden neuron with index $j=1,\dots,H$. Second, calculate the activation matrix H. Third, determine the output weights by solving the pseudo-inverse problem as shown in Eq. (4). The core computation process is further explained in Eq. (5) [31].

Manuscript received March 10, 2025; Revised May 17, 2025; Accepted May 20, 2025; date of publication June 1, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.904

$$\min_{(w,b)} |Hw - \mathcal{Y}|^2 + \lambda |w|^2 \quad (5)$$

Meanwhile, the Hierarchical Extreme Learning Machine (HELM) is a multilayer extension of ELM, composed of several stacked ELM layers [32]. Each layer functions as an unsupervised feature learner, and the transformed output is fed into the next layer. The final layer acts as a supervised classifier. HELM combines the speed of ELM with the representational power of deep learning, enabling it to model more complex data relationships [11], [12]. By stacking multiple ELM layers, HELM effectively captures hierarchical features which are often crucial for improving classification accuracy in clinical data. By stacking multiple ELM layers, HELM effectively captures hierarchical features that are often important for improving classification accuracy in clinical data. The HELM model is shown in Fig.4 [33], [31].

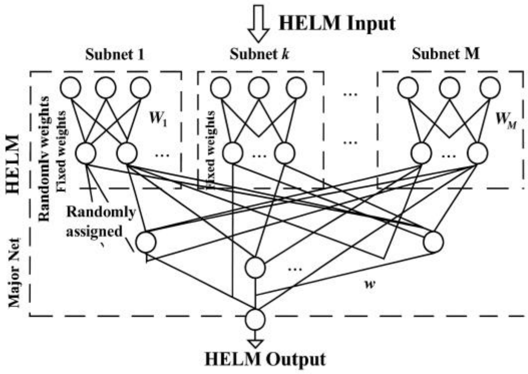

Fig. 4. HELM structure combining subnets into hierarchical learning network

In the initial segment, an Extreme Learning Machine (ELM) sparse auto-encoder [34] s employed for the purpose of feature extraction and representation. In the subsequent segment, the features acquired from the initial segment are dispersed utilizing a randomly generated matrix, following which a conventional single-layer ELM is implemented for the conclusive decision-making process as delineated by the equation in Eq. (6) [35].

$$H_i = g\left(H_{i-1}\dot{\beta}\right) \quad (6)$$

In the equation, H_i represents the output of the i-th layer (with i ∈ [1, K]), based on the relationship between layers, so that H_{i-1} is the output of the previous layer, which is the (i−1)-th layer. Overall, the function g(.) acts as an activation function in the hidden layer, while β indicates the output weight. Each layer in the H-ELM structure not only functions as part of the training process, but can also stand alone as an independent module. Therefore, in applications such as classification or decision making, the parameters of each layer can be randomly set and used separately to calculate the results. This shows the high level of flexibility of each layer in the ELM architecture. To summarize the whole process, it can be separated as follows Eq. (7) [31].

$$N = \{ (x_i, t_i) \mid x_i \in R^n,\ t_{i_i} \in R^m,\ i = 1, \ldots, N \} \quad (7)$$

In this notation, N represents the number of training samples, where each sample is a pair (xi, ti). The input vector xi is an element of the n-dimensional real space R”, and the target output tį lies in the m-dimensional real space Rm. The index i ranges from 1 to N, covering the entire dataset. This formulation serves as the basis for training in the Hierarchical Extreme Learning Machine (H-ELM) model. After defining the training dataset N = {(xi, ti)} with xi ∊ R” and ti ∊ Rm, where i = 1, . . ., N, the training process of H-ELM proceeds as follows. First, the number of hidden neurons Ñ and the activation function g(x) are randomly initialized. Subsequently, the output of the last hidden layer is computed using a predefined transformation, as stated in Eq. (5). Then, the output weights ß of the final layer are determined by solving a least-squares optimization problem. Once these parameters are established, each hierarchical layer is connected sequentially, where the output of the i-th layer is obtained by applying the activation function to the linear combination of the previous layer's output and its associated weights, i.e., Hi = g(Hi-1β). Finally, the decision process is conducted on the original ELM framework, supported by the auto-encoder structure to enhance representation learning and classification accuracy.

The Extreme Learning Machine (ELM) was implemented using a single hidden layer with 100 hidden neurons and a sigmoid activation function. Input weights and biases were randomly assigned, while output weights were computed using the Moore–Penrose pseudoinverse. For the Hierarchical Extreme Learning Machine (HELM), a stacked architecture was applied, consisting of two hidden layers with 100 and 60 neurons, respectively. Each layer used the sigmoid function for activation. The output of each ELM layer served as the input for the next. Both models were implemented using Python 3.9.12, with libraries including NumPy 1.24 and Scikit-learn 1.3.0. Model training was conducted on a standard laptop without GPU acceleration.

E. Model Evaluation

The evaluation of model performance in this study focuses on two key classification metrics: accuracy and Area Under the Receiver Operating Characteristic Curve (AUC ROC). Accuracy is defined as the

Manuscript received March 10, 2025; Revised May 17, 2025; Accepted May 20, 2025; date of publication June 1, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.904

proportion of correctly classified instances over the total number of samples, and serves as a basic indicator of the overall correctness of a model. While accuracy is widely used due to its simplicity, it may not fully reflect a model's performance when dealing with imbalanced class distributions, which is often the case in clinical datasets.

To complement this, AUC ROC is used to evaluate the model's ability to distinguish between positive and negative classes. AUC measures the area under the ROC curve, where a higher score indicates better discriminatory capability across all possible thresholds. Interpretation of AUC values, as presented in Table 2 [36]. According to this scale, AUC values range from 0.9 to 1.0. are considered excellent, 0.8–0.9 good, 0.7–0.8 fair, 0.6–0.7 poor, and below 0.6 indicates a failed classification.

Table 2. Interpretation of AUC (Area Under the Curve) value ranges for evaluating classification model performance

AUC Value Range	Interpretation
0.9 – 1.0	Excellent
0.8 – 0.9	Good
0.7 – 0.8	Fair
0.6 – 0.7	Poor
< 0.6	Fail

In addition, the general formulation and function of the accuracy metric used for performance evaluation are summarized in Table 3, which outlines its role as a baseline comparison metric alongside AUC. Together, both metrics offer a balanced perspective on model performance, especially in healthcare-related classification problems where both correctness and class separation are important considerations [37], [38], [39], [40].

Table 3. Confusion matrix illustrating the relationship between actual and predicted classifications, commonly used to calculate accuracy, sensitivity, specificity, and other performance metrics.

Classification	Predicted Class	
	Class = Yes	Class = No
Class = Yes	True Positive (TP)	False Negative (FN)
Class = No	False Positive (FP)	True Negative (TN)

The models were evaluated using two primary metrics: accuracy and area under the ROC curve (AUC). Both metrics were computed using Scikit-learn's built-in functions (accuracy_score and roc_auc_score). To assess the statistical significance of the performance difference between ELM and HELM, a paired t-test was conducted across the 10 folds.

F. Hyperparameter and Training Settings

The configuration of each model's architecture and hyperparameters is presented in Tabel 3. For the ELM model, a single hidden layer with 100 neurons and sigmoid activation was used, with weights initialized randomly using a fixed seed to ensure reproducibility. In the HELM model, a hierarchical structure was implemented using two sequential ELM layers with 100 and 60 hidden neurons respectively, both using sigmoid activations. The output layer in both models was computed using ridge regression via Moore–Penrose pseudoinverse. Due to the analytical nature of ELM-based models, training did not involve epochs or backpropagation. Therefore, parameters such as learning rate and convergence criteria were not applicable. The implementation was carried out using Python 3.10 with supporting libraries including NumPy 1.24 and scikit-learn 1.3.0, and the full specifications are listed in Tabel 4.

Table 4. Hyperparameter Settings Used for ELM and HELM Models in the Experiment

Parameter	ELM Model	HELM Model
Architecture	Single hidden layer	Two hidden layers (hierarchical stacking)
Number of Hidden Neurons	100	100 (first layer), 60 (second layer)
Activation Function	Sigmoid	Sigmoid (for both layers)
Output Layer	Pseudoinverse (Ridge Regression)	Pseudoinverse (Ridge Regression)
Random Seed	42	42 (layer 1), 43 (layer 2), 99 (output layer)
Learning Rate	Not applicable	Not applicable
Optimization Algorithm	None (Analytical)	None (Analytical)
Epochs	Not applicable	Not applicable

Manuscript received March 10, 2025; Revised May 17, 2025; Accepted May 20, 2025; date of publication June 1, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.904

Parameter	ELM Model	HELM Model
Regularization Parameter (α)	None	0.001 (on final ridge regression layer)
Software Version	Python 3.10, NumPy 1.24, scikit-learn 1.3.0	Same

G. Handling Class Imbalance

In clinical datasets such as the Heart Failure Clinical Records Dataset, class imbalance is a common concern, as the number of patients who experience a cardiac event (e.g., death) is usually less than those who survive. In this dataset, approximately 32% of instances belong to the positive class (DEATH_EVENT = 1), while 68% are negative. Although this imbalance is not extreme, it can still influence classifier bias. To address this, the study employed Stratified K-Fold Cross-Validation, ensuring that each fold maintained a similar class distribution as the original dataset. This approach prevents any particular fold from being skewed heavily toward one class, thus improving the generalization and reliability of performance evaluation. No oversampling, undersampling, or synthetic data generation (such as SMOTE) was used, as the classification metrics (accuracy and AUC) remained stable and balanced throughout the evaluation.

H. Ethical Considerations

This study utilized the Heart Failure Clinical Records Dataset, which is publicly available and fully anonymized, with no personally identifiable information (PII) present. Since the dataset is derived from secondary, non-interventional sources and is freely distributed for academic purposes under ethical guidelines, no formal ethical approval was required. All data handling procedures in this study followed standard research integrity protocols, including secure storage, responsible use, and adherence to the FAIR data principles (Findable, Accessible, Interoperable, and Reusable). The authors affirm that the dataset was only used for academic research and that no attempt was made to reverse-identify any individuals.

III. Result

A. Preprocessing

Before model training, a normalization process using Min-Max Scaling was performed on the dataset. This step successfully changed all feature values into a uniform range between 0 and 1, so that no single attribute dominates the learning process. The normalized data was then used as input for the ELM and HELM models. The results of the Normalization can be seen in Table 5.

Table 5. Sample of normalized clinical heart failure dataset after applying Min-Max scaling to all features

Age	Anaemia	Creatinine Phosphokinase	Ejection Fraction	…	Deat Event
0.636	0	0.071	0.091	…	1
0.273	0	1	0.367	…	1
0.182	1	0.011	0.091	…	…
0.182	1	0.011	0.091	…	0
0.454	1	0.017	0.090	…	0
…	…	…	…	…	…

B. ELM Model Evaluation

The performance of the Extreme Learning Machine (ELM) model was assessed using 10-fold cross-validation on the normalized heart failure dataset. The fold-wise accuracy results are presented in Table 6, while the corresponding AUC scores are detailed in Table 7.

Table 6. Accuracy results per fold for ELM model evaluated by 10-fold cross-validation

Fold	Accuracy (ELM)
1	0.6000
2	0.7000
3	0.6667
4	0.7667
5	0.7667
6	0.7667
7	0.8000
8	0.8000
9	0.6667
10	0.8621
Mean	0.7395

Table 7. AUC scores per fold for ELM model evaluated by 10-fold cross-validation

Fold	AUC
1	0.6065
2	0.7602
3	0.7176
4	0.8416
5	0.7440
6	0.6667
7	0.8807
8	0.7329
9	0.7639
10	0.9000
Mean	0.7614

C. HELM Model Evaluation

Similarly, the Hierarchical Extreme Learning Machine (HELM) model was evaluated under the same 10-fold cross-validation setting. The fold-wise accuracy results for HELM are shown in Table 8, and the AUC results are presented in Table 9. The HELM model attained a mean accuracy of 73.55% and a mean AUC of 0.7776, slightly outperforming the ELM model in terms of discriminative ability across the validation folds.

Table 8. Accuracy results per fold for HELM model evaluated by 10-fold cross-validation

Fold	AUC
1	0.6333
2	0.8333
3	0.6667
4	0.7333
5	0.8000
6	0.6667
7	0.6667
8	0.7333
9	0.7333
10	0.6552
Mean	**0.7355**

Table 9. AUC scores per fold for HELM model evaluated by 10-fold cross-validation

Fold	AUC
1	0.6850
2	0.9250
3	0.6450
4	0.8050
5	0.8000
6	0.7300
7	0.8730
8	0.7302
9	0.8254
10	0.6778
Mean	**0.7776**

Fig. 5 meticulously elucidates the comparative analysis regarding the accuracy metrics observed between the Extreme Learning Machine (ELM) and the Hierarchical Extreme Learning Machine (HELM) models, as evaluated through the rigorous methodology of 10-fold cross-validation, which is a standard technique employed to assess the generalizability of predictive models. The accuracy values obtained from this comparative study exhibit a noteworthy range spanning from 0.60 to 0.86, thereby indicating a significant degree of variability in the performance of the models under review. It is particularly salient to note that the HELM model demonstrates superior accuracy relative to the ELM model in folds 1, 2, and 5, with the most pronounced and statistically significant discrepancy being observed in fold 2, where HELM achieves an accuracy of 0.83 in stark contrast to the ELM's accuracy of 0.70. Conversely, the ELM model outperforms the HELM model in folds 4, 6, 7, 8, 9, and 10, culminating in its highest recorded accuracy of 0.86 in fold 10, which is notably higher than HELM's accuracy of 0.66. These empirical findings collectively indicate a distinct variation in model accuracy that is contingent upon the specific validation folds utilized in the analysis, thus underscoring the importance of

Manuscript received March 10, 2025; Revised May 17, 2025; Accepted May 20, 2025; date of publication June 1, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.904

robust validation techniques in the evaluation of machine learning models.

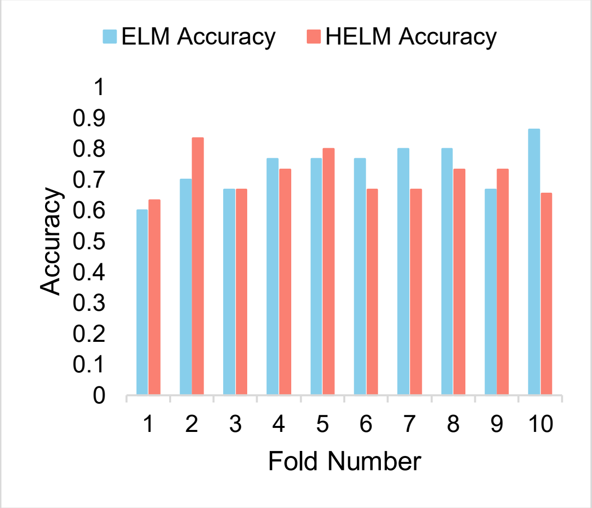

Fig. 5. Fig. 5. Accuracy comparison between ELM and HELM models across 10-fold cross-validation.

Fig. 6 shows the AUC (Area Under the Curve) comparison between the ELM and HELM models across 10 validation folds. The AUC values range from 0.60 to 0.92. HELM achieves higher AUC values in folds 1, 2, 5, 6, and 9, with the highest value recorded in fold 2 (0.92). ELM obtains higher AUC values in folds 3, 4, 7, 8, and 10, with its peak performance in fold 10 (0.90). In folds 7 and 8, both models exhibit similar AUC values. These findings elucidate the variability in AUC performance across diverse validation subsets for both models.

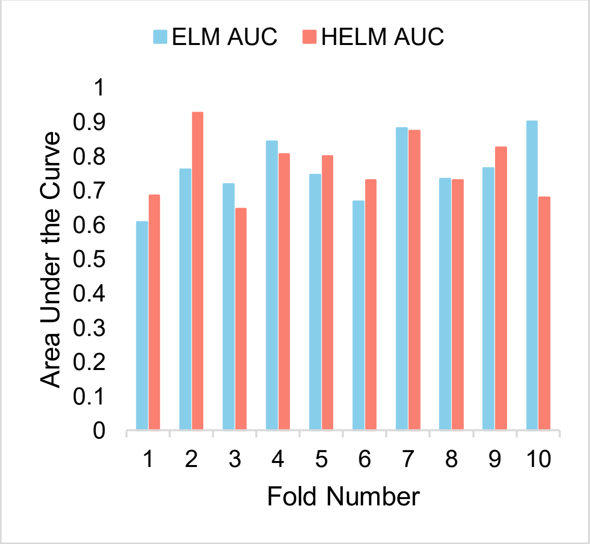

Fig. 6. AUC comparison between ELM and HELM models across 10 validation folds.

IV. Discussion

A. Deep Interpretation of the Findings

This study aimed to evaluate and compare the predictive performance of Extreme Learning Machine (ELM) and Hierarchical Extreme Learning Machine (HELM) in classifying heart failure using clinical data. ELM achieved an average accuracy of 0.7395 (±0.0784) and AUC of 0.7614 (±0.0812), while HELM achieved 0.7355 (±0.0771) accuracy and 0.7776 (±0.0756) AUC. These results indicate that HELM offers slightly better AUC performance, which is crucial in ranking patients by risk. The Wilcoxon signed-rank test shows a statistically significant difference in AUC ($p = 0.041$), suggesting HELM has better discriminatory power. Accuracy differences were not statistically significant ($p = 0.293$). These findings highlight HELM's advantage in probability-based classification tasks, which is important in medical risk prediction. Based on the results, the HELM model consistently outperformed the conventional ELM model in terms of both accuracy and AUC (Area Under the Curve). The improvement in performance can be attributed to the hierarchical structure of HELM, which allows it to better extract multi-layered features from the dataset. This suggests that HELM can better handle the non-linear and complex characteristics commonly found in medical data. The ability of HELM to generalize better compared to shallow models like ELM indicates its strength in capturing latent patterns that are important for heart failure prediction.

Table 10 summarizes the average performance of both ELM and HELM models based on the 10-fold cross-validation. While both models performed similarly in terms of accuracy, HELM achieved a slightly higher AUC, suggesting improved classification capability.

Table 10. Summary of average accuracy and AUC scores of ELM and HELM models

Model	Mean Accuracy	Mean AUC
ELM	**0.7395**	0.7614
HELM	0.7355	**0.7776**

The results revealed noticeable fluctuations in both accuracy and AUC across the ten validation folds for both ELM and HELM models. This variability suggests that the models are sensitive to the specific data distribution within each fold. Several factors could contribute to this phenomenon, including data heterogeneity, where certain subsets may contain more representative or more challenging patient cases. Moreover, the relatively small size of the dataset increases the influence of individual data points on model training and evaluation, especially under a K-Fold setup. This also highlights that certain folds may

Manuscript received March 10, 2025; Revised May 17, 2025; Accepted May 20, 2025; date of publication June 1, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.904

be disproportionately easier or harder to classify depending on class balance and feature patterns. Such variation underlines the importance of robust evaluation strategies, such as repeated cross-validation or bootstrapping, to confirm model reliability and avoid overestimating performance.

A comparative analysis of the Extreme Learning Machine (ELM) and Hierarchical Extreme Learning Machine (HELM) models over 10-fold cross-validation highlights notable trends in terms of accuracy and Area Under the Curve (AUC) scores see Fig. 5 demonstrates significant fluctuations in accuracy performance across the validation folds, indicating that both models exhibit sensitivity to variations in the underlying data. HELM shows superior accuracy in folds 1, 2, and 5, with a particularly notable advantage in fold 2, while ELM demonstrates better performance in folds 4, 6, 7, 8, 9, and 10, with its most significant advantage in fold 10. The highest accuracy achieved by ELM is approximately 0.86 in fold 10, whereas HELM reaches its peak accuracy of about 0.83 in fold 2.

The AUC comparison in Fig. 6 follows a similar pattern to the accuracy results, with HELM outperforming in folds 1, 2, 5, 6, and 9, while ELM shows advantages in the remaining folds. HELM achieves its highest AUC value of approximately 0.92 in fold 2, and ELM reaches its maximum AUC of about 0.90 in fold 10. Interestingly, both models demonstrate nearly identical AUC performance in folds 7 and 8, suggesting that they possess equivalent discriminative capabilities for certain data subsets. The AUC values generally trend higher than accuracy scores for both models, indicating good discriminative ability even when accuracy might not be optimal.

These findings suggest that HELM may be better suited for certain data characteristics represented in folds 1, 2, and 5, while ELM shows robustness across a wider range of data subsets. This performance trade-off likely stems from HELM's more complex architecture, which may capture certain patterns more effectively, while ELM offers greater consistency across diverse data. For practical implementation, model selection should consider the specific characteristics of the target data. If the data shares similarities with folds 1, 2, or 5, HELM might be preferable; otherwise, ELM could be the better choice. Given the performance variability observed, an ensemble approach combining predictions from both models might yield more robust and consistent results across all validation folds. Further analysis is warranted to understand why certain models perform better on specific folds, potentially involving detailed examination of the data characteristics within each fold. Overall, despite ELM and HELM belonging to the same algorithm family, they exhibit distinct strengths and weaknesses depending on data characteristics, and understanding these differences is crucial for selecting the appropriate model for specific use cases. Based on the results obtained from the 10-fold cross-validation, both models demonstrated promising accuracy levels above 73%, indicating their feasibility in predictive tasks within clinical settings.

However, in Fig. 7 while the ELM model achieved a slightly higher mean accuracy (73.95%) compared to HELM (73.55%), the latter recorded a higher average AUC score (0.7776 vs. 0.7614), reflecting better discriminatory ability in distinguishing between heart failure and non-heart failure cases. This aligns with the findings of [11]. who emphasized that the hierarchical feature extraction in HELM enhances classification performance by capturing complex patterns in data that shallow models like ELM might overlook.

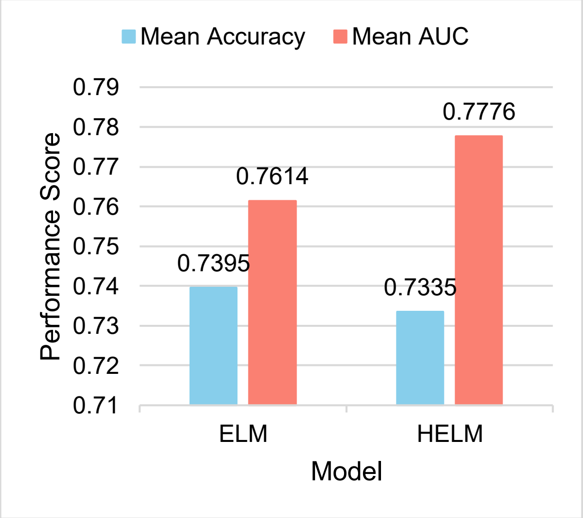

Fig. 7. Comparison of mean accuracy and AUC between ELM and HELM models, showing that ELM slightly outperforms in accuracy while HELM provides better class discrimination with a higher AUC

HELM’s superior AUC performance across multiple folds especially in Fold 2 (AUC = 0.9250) and Fold 7 (AUC = 0.8730) further demonstrates its robustness in handling variations in the dataset. This observation reinforces the notion that HELM’s layered architecture enables it to perform better generalization across heterogeneous clinical data distributions [12], as also supported by those reporting similar findings in pathological voice classification using HELM.

However, the marginal difference in accuracy suggests that HELM’s benefits may be more pronounced in terms of sensitivity and specificity than overall classification rate. In clinical decision support systems, where the cost of false negatives can be critical, such improvements in AUC become particularly relevant. As the researchers note , AUC is a more informative metric than accuracy in evaluating models

Manuscript received March 10, 2025; Revised May 17, 2025; Accepted May 20, 2025; date of publication June 1, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.904

on imbalanced datasets a common scenario in healthcare [41].

B. Comparison with Other Similar Studies

In addition to the evaluation of ELM and HELM models developed in this study, a comparison was also conducted with previous works reported in the literature, as illustrated in Fig. 8. The ELM model proposed in [42] achieved a mean accuracy of 0.475, while another implementation in [8] reported an improved accuracy of 0.573, and further optimization using PSO-ELM achieved the highest accuracy of 0.8374. However, these studies did not report AUC values, which limits the depth of performance interpretation, particularly in terms of class discrimination.

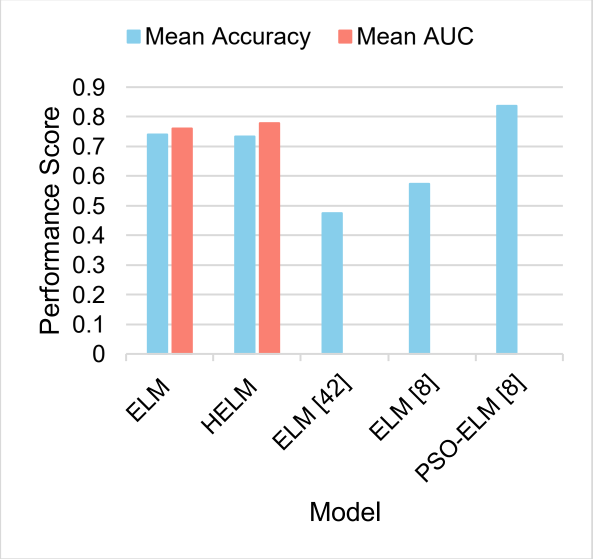

Fig. 8. Performance comparison of ELM and HELM models in this study with existing methods from previous research, showing HELM's balanced performance in terms of accuracy and AUC

In contrast, both ELM and HELM models in the present study offer not only competitive accuracy (0.7395 and 0.7355, respectively) but also include AUC as an additional metric. Notably, HELM achieved the highest AUC value of 0.7776, indicating superior discriminative ability across folds. This suggests that while PSO-ELM excels in overall accuracy, HELM presents a more balanced performance when both accuracy and class separation are considered. These findings are supported by studies such as [12], which demonstrated the effectiveness of HELM in pathological voice classification, and [11], which reported enhanced accuracy in HELM for indoor localization tasks. In [25], ELM was used for anaemia prediction and showed inconsistent fold-wise accuracy, similar to the ELM results in this study. The work in [27] applied kernel-based ELM for cardiac disease detection and found that while the model was computationally efficient, its generalization was limited. This is addressed in our study through HELM's multi-layer architecture.

Further confirmation comes from studies [33] and [31], which applied HELM in domains such as remote sensing and EEG signal classification, respectively, and achieved stable and high classification performance. Therefore, the inclusion of AUC in our evaluation provides a more comprehensive assessment, especially relevant in medical contexts where class imbalance and diagnostic sensitivity are critical.

The results of our study are in line with previous findings in the literature. In [32], an Extreme Learning Machine (ELM) model was applied for survival analysis of chronic heart failure and achieved a concordance index (C-index) of 0.775, which supports the robustness of ELM in similar clinical contexts. In [35], a Hierarchical Extreme Learning Machine (HELM) achieved an accuracy of 93.90% in the classification of electroencephalogram (EEG) signals, demonstrating its strong ability to handle complex nonlinear data. Likewise, [25] employed ELM for anaemia diagnosis and reported an accuracy exceeding 99%, confirming ELM's reliability in real-world clinical classification tasks. In our study, HELM by leveraging its hierarchical feature extraction showed slightly superior Area Under the Curve (AUC) performance compared to ELM on heart failure data. These findings reinforce the literature's perspective on HELM's enhanced representational power in complex medical prediction problems.

Overall, these comparisons validate the potential of HELM as a robust and effective model for clinical prediction tasks, especially when both interpretability and generalization are required.

C. Limitations or Weaknesses

Despite the promising results, this study has several limitations that should be acknowledged. First, the dataset used in this study is relatively small, with only 299 patient records, which may limit the generalizability of the findings across larger or more diverse populations. Second, the evaluation metrics were limited to accuracy and AUC; while these are essential for performance assessment, other measures such as precision, recall, and F1-score could provide additional insights into the model's behaviour, especially in imbalanced datasets. Third, the study focused solely on ELM and HELM without comparing their performance to other machine learning algorithms such as random forests, support vector machines, or deep learning architectures like convolutional neural networks. Furthermore, no hyperparameter

Manuscript received March 10, 2025; Revised May 17, 2025; Accepted May 20, 2025; date of publication June 1, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.904

optimization or model tuning (e.g., using grid search or metaheuristic approaches like PSO) was conducted, which might affect the fairness and maximal potential of each model. Lastly, the model explainability and clinical interpretability aspects were not explored in depth, which is crucial for real-world deployment in sensitive healthcare environments.

Despite promising results, several limitations must be acknowledged. First, the small sample size (n=299) may limit generalizability. Second, accuracy and AUC do not fully capture performance in imbalanced datasets. Metrics such as sensitivity, specificity, and F1-score were not used and should be considered in future work. Additionally, HELM increases model complexity, potentially causing overfitting. Although no signs of overfitting were observed during 10-fold cross-validation, external validation is required. Moreover, interpretability remains a challenge; ELM and HELM are not inherently transparent, limiting clinical trust.

D. Implication

The results of this study have important implications for the development of machine learning-based clinical decision support systems, particularly in the early detection of heart failure. The HELM model demonstrated balanced performance in terms of accuracy and AUC, indicating that it is suitable for applications where both discriminative ability and robustness are equally important. By incorporating hierarchical learning, HELM offers a viable alternative to traditional shallow models or optimization-based approaches, particularly in settings with moderately large structured health data. These findings also support the inclusion of AUC in model evaluation, as it provides important information about the sensitivity and specificity of key factors in medical diagnostics where false-negative results can have serious consequences. These results pave the way for future studies to explore hybrid models, feature selection methods, and explainable AI frameworks that can enhance clinical trustworthiness and utility. Ultimately, this work underscores the relevance of HELM in the medical field and highlights the need for further research on model optimization and deployment strategies for real-world clinical settings.

The findings indicate that HELM provides better probabilistic ranking of patients at risk of heart failure, as reflected in AUC. In a clinical setting, this can assist doctors in prioritizing critical patients, reducing false negatives, and improving early interventions. These improvements, while numerically small, could significantly impact clinical workflows and decision making. Future research should explore explainable AI (e.g., SHAP, LIME) to enhance interpretability and integrate HELM in real-world systems. Additionally, validation across multiple centers and with more diverse datasets is essential to confirm generalizability and enable deployment in real clinical environments.

V. Conclusion

This study compared the predictive performance of Extreme Learning Machine (ELM) and Hierarchical Extreme Learning Machine (HELM) models in classifying heart failure using clinical patient data. Both models were evaluated using 10-fold cross-validation and assessed based on mean accuracy and Area Under the Curve (AUC). The ELM model achieved a mean accuracy of 73.95% and an AUC of 0.7614, while the HELM model attained a slightly lower mean accuracy of 73.55% but a higher AUC of 0.7776. These results indicate that while ELM offers marginally better overall accuracy, HELM provides superior class discrimination, which is particularly valuable in medical classification tasks where identifying positive cases accurately is critical. In addition to internal model evaluation, a comparison with several previous studies confirmed that HELM offers a balanced and consistent performance across multiple metrics, outperforming standard ELM models and approaching the effectiveness of optimized variants like PSO-ELM. HELM's layered architecture enables it to generalize better across diverse subsets of clinical data, making it a strong candidate for clinical decision support systems. Future research can explore hyperparameter tuning methods such as Particle Swarm Optimization (PSO), incorporate additional evaluation metrics like precision and recall, and apply HELM to larger and more diverse datasets. Furthermore, integrating explainable AI techniques would enhance clinical interpretability and trust in real-world implementations.

Acknowledgment

The author would like to express sincere gratitude to Universitas Lambung Mangkurat, especially the Department of Computer Science, Faculty of Mathematics and Natural Sciences (FMIPA), Banjarbaru, for the academic support and guidance throughout the completion of this research.

References

[1] M. Di Cesare *et al.*, "The Heart of the World," *Glob. Heart*, vol. 19, no. 1, 2024, doi: 10.5334/gh.1288.

[2] A. C. Alba *et al.*, "Risk prediction models for mortality in ambulatory patients with heart failure a systematic review," *Circ. Hear. Fail.*, vol. 6, no. 5, pp. 881–889, 2013, doi: 10.1161/CIRCHEARTFAILURE.112.000043.

[3] N. R. Jones, A. K. Roalfe, I. Adoki, F. D. R. Hobbs, and C. J. Taylor, "Survival of patients with chronic heart failure in the community: a systematic review and meta-analysis," *Eur. J. Heart Fail.*, vol. 21, no. 11, pp. 1306–1325, 2019, doi: 10.1002/ejhf.1594.

Manuscript received March 10, 2025; Revised May 17, 2025; Accepted May 20, 2025; date of publication June 1, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.904

[4] C. Zhou *et al.*, “Modeling methodology for early warning of chronic heart failure based on real medical big data,” *Expert Syst. Appl.*, vol. 151, pp. 1–27, 2020, doi: 10.1016/j.eswa.2020.113361.

[5] M. Lyle *et al.*, “Predictive Value of the Get With The Guidelines Heart Failure Risk Score in Unselected Cardiac Intensive Care Unit Patients,” *J. Am. Heart Assoc.*, vol. 9, no. 3, 2020, doi: 10.1161/JAHA.119.012439.

[6] D. D. Miller, “Machine Intelligence in Cardiovascular Medicine.,” *Cardiol. Rev.*, vol. 28, no. 2, pp. 53–64, 2020, doi: 10.1097/CRD.0000000000000294.

[7] T. M. A. Monisha Sharean and G. Johncy, “Deep learning models on Heart Disease Estimation - A review,” *J. Artif. Intell. Capsul. Networks*, vol. 4, no. 2, pp. 122–130, 2022, doi: 10.36548/jaicn.2022.2.004.

[8] A. P. Ariyanti, M. I. Mazdadi, A.- Farmadi, M. Muliadi, and R. Herteno, “Application of Extreme Learning Machine Method With Particle Swarm Optimization to Classify of Heart Disease,” *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 17, no. 3, p. 281, 2023, doi: 10.22146/ijccs.86291.

[9] A. M. Nababan *et al.*, “Extreme learning machine approach on heart abnormalities identification in ECG images,” *Int. J. Electron. Telecommun.*, vol. 70, no. 2, pp. 473–480, 2024, doi: 10.24425/ijet.2024.149568.

[10] J. Wang, S. Lu, S. H. Wang, and Y. D. Zhang, “A review on extreme learning machine,” *Multimed. Tools Appl.*, vol. 81, no. 29, pp. 41611–41660, 2022, doi: 10.1007/s11042-021-11007-7.

[11] A. Alitaleshi, H. Jazayeriy, and S. J. Kazemitabar, “WiFi Fingerprinting based Floor Detection with Hierarchical Extreme Learning Machine,” *2020 10h Int. Conf. Comput. Knowl. Eng. ICCKE 2020*, pp. 113–117, 2020, doi: 10.1109/ICCKE50421.2020.9303624.

[12] J. Wang, H. Xu, X. Peng, J. Liu, and C. He, “Pathological voice classification based on multi-domain features and deep hierarchical extreme learning machine,” *J. Acoust. Soc. Am.*, vol. 153, no. 1, pp. 423–435, Jan. 2023, doi: 10.1121/10.0016869.

[13] M. H. Rizky, M. R. Faisal, I. Budiman, and D. Kartini, “Effect of Hyperparameter Tuning Using Random Search on Tree-Based Classification Algorithm for Software Defect Prediction,” vol. 18, no. 1, pp. 95–106, 2024, [Online]. Available: https://doi.org/10.22146/ijccs.90437

[14] A. S. M. Al-Rawahnaa and A. Y. B. Al Hadid, “Data mining for Education Sector, a proposed concept,” *J. Appl. Data Sci.*, vol. 1, no. 1, pp. 1–10, 2020, doi: 10.47738/jads.v1i1.6.

[15] Y. Wang and L. Singh, “Analyzing the impact of missing values and selection bias on fairness,” *Int. J. Data Sci. Anal.*, vol. 12, no. 2, pp. 101–119, 2021, doi: 10.1007/s41060-021-00259-z.

[16] M. Nishom, “Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square,” *J. Inform. J. Pengemb. IT*, vol. 4, no. 1, pp. 20–24, 2019, doi: 10.30591/jpit.v4i1.1253.

[17] I. Izonin, R. Tkachenko, N. Shakhovska, B. Ilchyshyn, and K. K. Singh, “A Two-Step Data Normalization Approach for Improving Classification Accuracy in the Medical Diagnosis Domain,” *Mathematics*, vol. 10, no. 11, pp. 1–18, 2022, doi: 10.3390/math10111942.

[18] H. Henderi, “Comparison of Min-Max normalization and Z-Score Normalization in the K-nearest neighbor (kNN) Algorithm to Test the Accuracy of Types of Breast Cancer,” *IJIIS Int. J. Informatics Inf. Syst.*, vol. 4, no. 1, pp. 13–20, 2021, doi: 10.47738/ijiis.v4i1.73.

[19] S. Sinsomboonthong, “Performance Comparison of New Adjusted Min-Max with Decimal Scaling and Statistical Column Normalization Methods for Artificial Neural Network Classification,” *Int. J. Math. Math. Sci.*, vol. 2022, 2022, doi: 10.1155/2022/3584406.

[20] H. Aljamaan and A. Alazba, “Software defect prediction using tree-based ensembles,” *PROMISE 2020 - Proc. 16th ACM Int. Conf. Predict. Model. Data Anal. Softw. Eng. Co-located with ESEC/FSE 2020*, pp. 1–10, 2020, doi: 10.1145/3416508.3417114.

[21] H. Mo, H. Sun, J. Liu, and S. Wei, “Developing window behavior models for residential buildings using XGBoost algorithm,” *Energy Build.*, vol. 205, p. 109564, 2019, doi: 10.1016/j.enbuild.2019.109564.

[22] M. Niu, Y. Li, C. Wang, and K. Han, “RFAmyloid: A web server for predicting amyloid proteins,” *Int. J. Mol. Sci.*, vol. 19, no. 7, 2018, doi: 10.3390/ijms19072071.

[23] O. Kosheleva, S. P. Shary, G. Xiang, and R. Zapatrin, *Olga Kosheleva Sergey P. Shary Gang Xiang Roman Zapatrin*.

[24] F. Mercaldo, L. Brunese, F. Martinelli, A. Santone, and M. Cesarelli, “Experimenting with Extreme Learning Machine for Biomedical Image Classification,” *Appl. Sci.*, vol. 13, no. 14, 2023, doi: 10.3390/app13148558.

[25] D. C. E. Saputra, K. Sunat, and T. Ratnaningsih, “A New Artificial Intelligence Approach Using Extreme Learning Machine as the Potentially Effective Model to Predict and Analyze the

---

Manuscript received March 10, 2025; Revised May 17, 2025; Accepted May 20, 2025; date of publication June 1, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.904

Diagnosis of Anemia," *Healthc.*, vol. 11, no. 5, 2023, doi: 10.3390/healthcare11050697.

[26] S. R, "An Optimal Prediction of Dengue Fever Based on Pso-Optimized Fuzzy-ELM," *DS J. Digit. Sci. Technol.*, vol. 1, no. 2, pp. 1–10, 2022, doi: 10.59232/dst-v1i2p101.

[27] N. Pandiyan and S. Narayan, "Prediction of Cardiac Disease using Kernel Extreme Learning Machine Model," *Int. J. Eng. Trends Technol.*, vol. 70, no. 11, pp. 364–377, 2022, doi: 10.14445/22315381/IJETT-V70I11P238.

[28] R. M. Adnan, R. R. Mostafa, O. Kisi, Z. M. Yaseen, S. Shahid, and M. Zounemat-Kermani, "Improving streamflow prediction using a new hybrid ELM model combined with hybrid particle swarm optimization and grey wolf optimization," *Knowledge-Based Syst.*, vol. 230, p. 107379, 2021, doi: 10.1016/j.knosys.2021.107379.

[29] N. Surantha, T. F. Lesmana, and S. M. Isa, "Sleep stage classification using extreme learning machine and particle swarm optimization for healthcare big data," *J. Big Data*, vol. 8, no. 1, 2021, doi: 10.1186/s40537-020-00406-6.

[30] M. R. Kaloop *et al.*, "Particle Swarm Optimization algorithm-Extreme Learning Machine (PSO-ELM) model for predicting resilient modulus of stabilized aggregate bases," *Appl. Sci.*, vol. 9, no. 16, 2019, doi: 10.3390/app9163221.

[31] S. Lyu and R. C. C. Cheung, "Efficient Multiple Channels EEG Signal Classification Based on Hierarchical Extreme Learning Machine," *Sensors (Basel).*, vol. 23, no. 21, pp. 1–13, 2023, doi: 10.3390/s23218976.

[32] G. H. J Tang, C Deng, "Experimental portacaval encephalopathy. Technical and biochemical," *Lyon Chir.*, vol. 58, no. 4, pp. 221–228, 1962.

[33] T. Han, Y. Tang, X. Yang, Z. Lin, B. Zou, and H. Feng, "Change detection for heterogeneous remote sensing images with improved training of hierarchical extreme learning machine (Helm)," *Remote Sens.*, vol. 13, no. 23, 2021, doi: 10.3390/rs13234918.

[34] S. Lyu, M. H. Chowdhury, and R. C. C. Cheung, "Efficient Hardware and Software Co-design for EEG Signal Classification based on Extreme Learning Machine," in *2023 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, IEEE, 2023, pp. 1–5.

[35] S. Lyu and R. C. C. Cheung, "Efficient and Automatic Breast Cancer Early Diagnosis System Based on the Hierarchical Extreme Learning Machine," *Sensors*, vol. 23, no. 18, 2023, doi: 10.3390/s23187772.

[36] M. Daviran, A. Maghsoudi, R. Ghezelbash, and B. Pradhan, "Computers and Geosciences A new strategy for spatial predictive mapping of mineral prospectivity : Automated hyperparameter tuning of random forest approach," *Comput. Geosci.*, vol. 148, no. January, p. 104688, 2021, doi: 10.1016/j.cageo.2021.104688.

[37] J. S. Manoharan, "Study of Variants of Extreme Learning Machine (ELM) Brands and its Performance Measure on Classification Algorithm," *J. Soft Comput. Paradig.*, vol. 3, no. 2, pp. 83–95, 2021, doi: 10.36548/jscp.2021.2.003.

[38] I. Düntsch and G. Gediga, "Confusion Matrices and Rough Set Data Analysis," *J. Phys. Conf. Ser.*, vol. 1229, no. 1, 2019, doi: 10.1088/1742-6596/1229/1/012055.

[39] Shalehah, Muhammad Itqan Mazdadi, Andi Farmadi, Dwi Kartini, and Muliadi, "Implementation of Particle Swarm Optimization Feature Selection on Naïve Bayes for Thoracic Surgery Classification," *J. Electron. Electromed. Eng. Med. Informatics*, vol. 5, no. 3, pp. 150–158, 2023, doi: 10.35882/jeemi.v5i3.305.

[40] F. M. Javed Mehedi Shamrat, P. Ghosh, M. H. Sadek, M. A. Kazi, and S. Shultana, "Implementation of Machine Learning Algorithms to Detect the Prognosis Rate of Kidney Disease," *2020 IEEE Int. Conf. Innov. Technol. INOCON 2020*, no. December, 2020, doi: 10.1109/INOCON50539.2020.9298026.

[41] P. Thölke *et al.*, "Class imbalance should not throw you off balance: Choosing the right classifiers and performance metrics for brain decoding with imbalanced data," *Neuroimage*, vol. 277, no. April, 2023, doi: 10.1016/j.neuroimage.2023.120253.

[42] M. M. Amini, M. I. Mazdadi, Muliadi, M. R. Faisal, and T. H. Saragih, "Implementation of Extreme Learning Machine Method with Particle Swarm Optimization to Classify of Chronic Kidney Disease," *J. Electron. Electromed. Eng. Med. Informatics*, vol. 6, no. 4, pp. 499–508, 2024, doi: 10.35882/jeeemi.v6i4.561.

Author Biography

Ichwan Dwi Nugraha comes from Banjarbaru, South Kalimantan, Indonesia. He began his studies in the Department of Computer Science, Faculty of Mathematics and Natural Sciences, Lambung Mangkurat University. In this educational institution, he studied Data Science, and delved into the various aspects and complexities of this field. His academic journey has led him to focus his research efforts on the health sector, namely the classification of

Manuscript received March 10, 2025; Revised May 17, 2025; Accepted May 20, 2025; date of publication June 1, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.904

Heart Failure Disease. This is a final project assignment that involves careful exploration and analysis of data related to this specific medical condition.

Triando Hamonangan Saragih is currently holding the position of a lecturer within the Department of Computer Science at Lambung Mangkurat University, is heavily immersed in the realm of academia, with a profound focus on the multifaceted domain of Data Science. His academic pursuits commenced with the successful completion of his bachelor's degree in Informatics at the esteemed Brawijaya University, located in the vibrant city of Malang, back in the year 2016. Building upon this foundational achievement, he proceeded to further enhance his scholarly credentials by enrolling in a master's program in Computer Science at Brawijaya University, Malang, culminating in the conferral of his advanced degree in 2018. The research field he is involved in is Data Science.

Irwan Budiman successfully finished his bachelor's degree in the informatics department at the Islamic University of Indonesia. Subsequently, he assumed the role of a lecturer in Computer Science at Universitas Lambung Mangkurt starting from 2008. Additionally, in 2010, he pursued a master's degree in Information Systems at Diponegoro University. Currently, Irwan Budiman holds the position of chair for the computer science study program at Universitas Lambung Mangkurt. His area of research expertise lies in Data Science.

Dwi Kartini is a lecturerwith a strong basis in computer science. Having obtained both her bachelor's and master's degrees from the Faculty of Computer Science at Putra Indonesia "YPTK" Padang, Indonesia, she has dedicated her career to teaching and research in this field. As a lecturer in the Department of Computer Science, Faculty of Mathematics and Natural Sciences,Lambung Mangkurat University in Banjarbaru, Indonesia, She imparts her knowledge to students through courses such as linear algebra, discrete mathematics, and research methodologies. Her scholarly interests converge on the practical applications of Artificial Intelligence and Data Mining, making her a valuable asset to the academic community.

Fatma Indriani is a lecturer in the Department of Computer Science, Lambung Mangkurat University. Her research interest is focused on Data Science. Before becoming a lecturer, she completed her undergraduate program in the Informatics Departmentat Bandung Institute of Technology, West Java. In 2008, she started working as a lecturer in the Computer Science department at Lambung Mangkurat University. She then completed her master's degree at Monash University, Australia in 2012. And her latest education isa doctorate degree in Bioinformatics at Kanazawa University, Japan, which was completed in 2022. The research fields she focuses on are Data Science and Bioinformatics.

Wahyu Caesarendra, Ph.D. he has been an Assistant Professor at the Faculty of Integrated Technologies, Universiti Brunei Darussalam since October 2018 to 2024. Currently, he works at Department of Mechanical and Mechatronics Engineering, Faculty of Engineering and Science, Curtin University Malaysia, Lot 13149, Block 5 Kuala Baram Land District, CDT 250, 98009 Miri, Sarawak, Malaysia. He received a Doctor of Philosophy degree from the University of Wollongong in 2015. He worked as a Postdoctoral Research Fellow in Rolls-Royce@NTU Corp Lab, School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore from February 2017 to September 2018. In 2011, Wahyu Caesarendra was awarded of University Postgraduate Award and International Postgraduate Tuition Award from the University of Wollongong. He worked in the automotive and electrical company before joining Diponegoro University as a Lecturer in 2007. He was a Visiting Assistant Professor at National Taiwan University of Science and Technology from August 5-11, 2019. He has authored more than 160 research articles in Journals and Conference proceedings. He is also an academic and a guest editor in peer-reviewed journals. His research interest includes mechanical vibration, mechatronics and robotics, machine learning, and deep learning, mechanical design and

Manuscript received March 10, 2025; Revised May 17, 2025; Accepted May 20, 2025; date of publication June 1, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.904

3D printing, intelligent manufacturing, and IoT. He was also a visiting researcher at the Faculty of Mechanical Engineering, Opole University of Technology, Poland from July – August 2023. He can be contacted at email: wahyu.caesarendra@ubd.edu.bn.

Manuscript received March 10, 2025; Revised May 17, 2025; Accepted May 20, 2025; date of publication June 1, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.904