PeerJ Computer Science

Prediction of body fat percentage using a new hybrid intelligent feature selection: ANN-KGA & ANN-IKGA

Tohid Yousefi[1] and Özlem Varlıklar[2]

[1] Faculty of Computer and Information Technologies, Cappadocia University, Urgup, Nevsehir, Turkey
[2] Department of Computer Engineering, Dokuz Eylül University, Izmir, Turkey

ABSTRACT

Before initiating obesity treatment, it is essential to determine body fat percentage (BFP), a measure that cannot be assessed through conventional weighing alone. To address this, specialized "Body Analyzers" have been developed; however, their high cost encourages the search for more practical and cost-effective alternatives. In this study, 16 machine learning algorithms (six linear and 10 non-linear) were evaluated for BFP prediction and compared with two hybrid evolutionary frameworks combining an Artificial Neural Network (ANN) with a K-means Genetic Algorithm (ANN-KGA) and an Improved K-means Genetic Algorithm (ANN-IKGA). To ensure methodological rigor and prevent optimistic bias, a strict nested cross-validation strategy (five outer folds × three inner folds) was employed during feature selection and model evaluation. To account for the stochastic nature of the genetic algorithm, the entire nested cross-validation procedure was repeated 15 times with different random seeds. The ANN-KGA model achieved a mean outer-fold R^2 of 0.7522 (95% CI [0.6963–0.8081]) with a mean Root Mean Squared Error (RMSE) of 3.9651 (95% CI [3.5453–4.3849]) across the 15 runs. The proposed ANN-IKGA model demonstrated strong predictive performance, achieving a mean outer-fold R^2 of 0.7858 (95% CI [0.7495–0.8222]) and a mean RMSE of 3.7182 (95% CI [3.1777–4.2586]) across the 15 independent runs. Statistical comparison against baseline models using paired t-tests indicated significant improvement ($p < 0.05$) for both evolutionary approaches. These findings demonstrate that the proposed ANN-IKGA framework enhances predictive accuracy and generalization performance under strict nested validation. The results highlight the effectiveness of evolutionary feature optimization in improving BFP estimation using anthropometric measurements, providing a reliable and cost-effective alternative to specialized body analysis equipment.

Subjects Artificial Intelligence, Computer Education, Data Mining and Machine Learning, Data Science

Keywords Artificial neural network, Body fat percentage, Body fat prediction, Data processing, Feature selection, Improved K-means genetic algorithm, K-means genetic algorithm, Machine learning, Meta-heuristic algorithms

INTRODUCTION

Each year, 17.9 million lives succumb to cardiovascular disease, surpassing the mortality rate of cancer (*World Health Organization, 2018*). Obesity, a significant risk factor for cardiovascular diseases, further compounds this health challenge (*Gruzdeva et al., 2018*).

Submitted 5 September 2025
Accepted 15 May 2026
Published 2 July 2026

Corresponding author
Tohid Yousefi,
tohid.yousefi@kapadokya.edu.tr

Academic editor
Luigi Di Biasi

Additional Information and Declarations can be found on page 34

DOI **10.7717/peerj-cs.3984**

The resulting interconnected ailments notably diminish the quality of life-related to health. Beyond mere statistical figures, the repercussions extend into daily experiences. Effectively addressing this issue requires holistic strategies implemented by healthcare professionals, policymakers, and society at large, all working collaboratively toward a healthier future (*Kopelman, 2007*).

In addition to causing a decline in productivity, certain diseases result in considerable healthcare expenses for those affected (*Wang et al., 2011*). The presence of excess body fat contributes to obesity, but low body fat can lead to issues like pubertal delay and osteoporosis in adolescents (*Gjesdal et al., 2008*). Assessing body fat percentage (BFP) by dividing body fat mass by body weight provides crucial insights into an individual's health status. Two main methods, direct and indirect, are employed to measure body fat. Direct methods, suitable for cadavers, include precise techniques like dual-energy X-ray absorptiometry but is expensive and not widely used clinically (*Ellis, 2000*). In contrast, anthropometry, a simple and cost-effective indirect method, uses indicators like body mass index and waist circumference to assess obesity-related health risks (*Imai et al., 2012*).

Numerous studies have focused on predicting body fat through regression, employing statistical methods to formulate linear or nonlinear equations based on anthropometric measurements (*Freedman et al., 2009*; *Johnson, 1996*; *Svendsen et al., 1991*). Furthermore, some studies have attempted to estimate body fat using fewer features through feature selection methods (*Chiong et al., 2021*; *Lai et al., 2022*; *Shao, 2014*; *Uçar et al., 2021*). In contrast, our novel approach in this study not only reduced the number of features but also improved body fat prediction performance, compared to previously reported approaches under a strictly controlled validation framework.

Feature selection constitutes a vital procedure in data pre-processing and holds significant importance in machine learning (*Kalousis, Prados & Hilario, 2007*). With the expansion of research in data mining, attention to feature selection has notably increased (*Mlambo, Cheruiyot & Kimwele, 2016*). This approach efficiently lowers data dimensionality, reduces computational expenses, and enhances classification performance (*Hua, Tembe & Dougherty, 2009*). Within data mining, feature selection represents a key stage that improves algorithmic efficiency by minimizing dataset size. Nevertheless, feature selection algorithms by themselves may not reach the optimal solution due to their restricted search capabilities (*Yusta, 2009*). Although heuristic techniques frequently yield acceptable solutions, they might be inadequate for complex challenges and heavily depend on the developer's expertise (*Chinneck, 2006*; *Coello, 2007*). In contrast, metaheuristic algorithms, which employ randomization alongside local search strategies, generally surpass simple heuristics (*Yang, 2010*). Metaheuristics have emerged as a widely adopted approach owing to their capacity for comprehensive exploration and effective resolution of intricate problems (*Glover & Kochenberger, 2006*). Therefore, metaheuristic approaches present a promising avenue for precisely tackling feature selection issues.

As previously noted, feature selection represents a critical stage in data preprocessing, enabling the reduction of data dimensionality and the enhancement of predictive performance. Nevertheless, conventional feature selection techniques often fail to deliver adequate solutions. To overcome this limitation, this study proposes a hybrid

methodology. In our research, all data preprocessing steps—except feature selection—were first conducted on the body fat dataset using 16 machine learning algorithms. Thereafter, we apply our proposed methods, which involve intelligent feature selection through the combination of an artificial neural network algorithm with both a K-means Genetic Algorithm and an Improved K-means Genetic Algorithm, to the same dataset. The results highlight the positive impact of utilizing hybrid meta-heuristic methods that integrate clustering with evolutionary search in feature selection, as it leads to enhanced performance of the final model.

Despite the growing body of research on feature selection for body fat prediction, many studies rely on single hold-out validation strategies, which may lead to optimistic bias when feature selection and model evaluation are not strictly separated. To address this methodological concern, the present study adopts a nested cross-validation framework that ensures complete separation between feature optimization and final performance evaluation (*Parvandeh et al., 2020*). This design enhances the reliability and reproducibility of the reported results.

The article introduces several significant contributions and innovations, which can be outlined as follows:

i. **Introduction of ANN-KGA and ANN-IKGA:** This article introduces two novel hybrid methods, specifically designed for intelligent feature selection.

ii. **Improved Results in Feature Selection:** The proposed methods outperform conventional machine learning techniques in the field of feature selection, yielding more promising and strong results.

iii. **Comprehensive Intelligent Methods:** The article presents ANN-KGA and ANN-IKGA as comprehensive intelligent methods based on artificial intelligence. These approaches offer effective solutions for a wide range of feature selection problems by balancing exploration and exploitation.

iv. **Selective Feature Utilization:** Through intelligent selection, the proposed ANN-IKGA method efficiently identifies and utilizes a small percentage of the most relevant features from the dataset, surpassing the performance of existing methods that utilize all features.

v. **Improved Predictive Performance (R^2):** In estimating body fat percentage, the proposed methods demonstrate competitive and improved performance relative to 16 machine learning algorithms, highlighting their effectiveness in producing reliable and accurate predictions.

vi. **Contextual Comparison with Literature:** Contextual Comparison with Literature: When considered in the context of previously reported results on the same dataset, the proposed method demonstrates performance levels that appear competitive with existing approaches. However, it is important to note that these comparisons are indicative rather than direct, as differences in preprocessing pipelines, feature selection strategies, and validation protocols across studies may affect the reported results (*Chiong et al., 2021*; *Lai et al., 2022*; *Shao, 2014*; *Uçar et al., 2021*).

vii. **Rigorous Validation Framework:** All performance evaluations reported in this study are obtained under a strict nested cross-validation procedure (five outer folds × three inner folds), repeated 15 times with different random seeds to account for stochastic variability. This ensures that feature selection, model training, and testing are completely isolated and that results are robust and reproducible.

In this study, the proposed model is intended as a low-cost and non-invasive screening and monitoring tool for estimating body fat percentage in adult populations. It is not designed to replace gold-standard diagnostic methods such as dual-energy X-ray absorptiometry (DXA) but rather to provide a practical alternative in primary care, community health programs, and large-scale epidemiological assessments where access to specialized equipment is limited. The primary objective is to support early risk identification and population-level health monitoring rather than definitive clinical diagnosis.

The article is organized as follows: In 'Related Works', related work on feature selection techniques is reviewed. 'Material and Methods Data Preprocessing', a comprehensive description of the materials and methods employed is provided, including an in-depth explanation of the proposed ANN-KGA and ANN-IKGA hybrid method for intelligent feature selection. 'Results' is dedicated to presenting the obtained results and conducting comparisons between the performance of 16 machine learning methods and the proposed hybrid method. Lastly, the article concludes with a section encompassing the key findings, conclusions, and research perspectives for future endeavors.

RELATED WORKS

Estimating body fat percentage from simple anthropometric or clinical measurements has become an active research area, as low-cost and accurate models are valuable for both clinical practice and large-scale studies. Existing work generally falls into three groups: (1) compact statistical formulas calibrated against reference methods, (2) studies that enhance inputs through feature engineering, and (3) hybrid or metaheuristic feature-selection frameworks integrated with machine-learning regressors. Below, representative studies from these directions are summarized, with emphasis on those most relevant to hybrid intelligent feature-selection methods.

Lai et al. (2022) propose a hybrid feature-selection framework that combines a VIKOR-based multi-filter ensemble with an improved simplified swarm optimization (iSSO) and a local-refinement step to predict body fat percentage. The approach first ranks candidate features using multiple filter criteria, then applies iSSO to search compact subsets and refines promising solutions locally, aiming to reduce prediction error while keeping model complexity low. Their experiments on body-composition datasets show that the hybrid pipeline reduces error *vs*. several baseline selection strategies and produces smaller feature subsets without sacrificing predictive performance (*Lai et al., 2022*).

Fan et al. (2022) focus on feature-extraction rather than selection. They evaluate multiple unsupervised transformations (including PCA and related approaches) applied to anthropometric and laboratory measurements and then train standard regressors on

the transformed representations. Their results on two public body-fat datasets indicate that transformed, low-dimensional feature sets often yield lower MAE and RMSE and increased robustness to noise and correlated inputs compared with using raw measurements. The article highlights when dimensionality reduction can be preferable to naïve selection, especially for noisy or highly multicollinear predictors (*Fan et al., 2022*).

Hussain, Cavus & Sekeroglu (2021) present a hybrid modeling approach (rather than a feature-selection method) that ensembles Support Vector Regression (SVR) with an emotional artificial neural network (EANN). The SVR–EANN pipeline blends complementary learners so that nonlinear relationships captured by the neural model complement the regularization and generalization strengths of SVR. The authors report improved predictive accuracy over single models and provide details on preprocessing and validation protocols that are useful when comparing hybrid feature-selection pipelines with hybrid model ensembles (*Hussain, Cavus & Sekeroglu, 2021*).

Keivanian, Chiong & Fan (2023) introduce a fuzzy adaptive evolutionary framework designed specifically for single- and multi-objective feature selection in body-fat prediction. Their method integrates fuzzy set theory with adaptive evolutionary operators to combine multiple selection criteria (accuracy, sparsity, robustness) through a fuzzy weighted sum and uses global/local evolutionary search to avoid local optima. The framework explicitly models trade-offs between conflicting objectives and demonstrates competitive predictive performance while giving users control over the balance between compactness and accuracy (*Keivanian, Chiong & Fan, 2023*).

Xu et al. (2023) take a complementary, parsimony-driven perspective: they develop and validate prediction equations for body-fat percentage obtained from DXA using simple predictors such as BMI and socio-demographic variables. Their calibrated regression equations are interpretable, cheap to compute, and perform robustly at the population level when validated against DXA. This study is a useful baseline and a reminder that, in some contexts, simple, well-validated equations can be preferable to complex machine-learning pipelines especially when interpretability and wide applicability are priorities (*Xu et al., 2023*).

Santos (2023) (preprint and public replications) evaluate standard regression and regularized baselines (ordinary least squares, ridge, Bayesian ridge) on body-fat datasets and argue that well-regularized linear models are strong, reproducible baselines when preprocessing and outlier handling are performed carefully. Their findings emphasize the importance of including simple, interpretable baselines when reporting gains from complex feature-selection or hybrid approaches (*Santos, 2023*).

Although numerous metaheuristic and multi-objective feature-selection methods exist in the literature, many are difficult to tune and apply in practice. To address this, our study introduces a hybrid intelligent approach—ANN-KGA and ANN-IKGA—for accurate and efficient body fat percentage prediction. By combining artificial neural networks with the K-means Genetic Algorithm and its improved variant, the method unifies data-driven learning and clustering-enhanced evolutionary optimization. This hybrid design offers a practical and high-performing alternative to purely optimization-focused techniques.

A brief comparison of key feature-selection methods and their characteristics is summarized in Table 1.

MATERIALS AND METHODS

Data preprocessing

Data mining aims to extract meaningful patterns from datasets; however, model success strongly depends on data quality. Raw data may contain missing, noisy, inconsistent, or extreme values, which can negatively affect predictive performance. Therefore, an effective data pre-processing stage is essential before developing the proposed ANN-KGA and ANN-IKGA models (*Zhang, Zhang & Yang, 2003*). In this study, to prevent any potential data leakage and ensure a fair evaluation, all preprocessing steps were integrated into the nested cross-validation framework. Specifically, data cleaning, outlier handling, feature engineering, and data transformation procedures were performed exclusively within each outer training fold, and the learned parameters were subsequently applied to the corresponding validation/test folds. This strategy guarantees a strict separation between training and testing data, ensuring that no information from the outer test folds is used during model training. The overall workflow (Fig. 1) includes data cleaning, data reduction, data transformation, and data scaling, all conducted in a fold-wise manner within the nested cross-validation process.

Data cleaning

Data cleaning is the first and most critical step in the preprocessing pipeline. Its purpose is to improve data quality by addressing missing values and outliers, thereby preventing biased or unstable ANN training (*Maingi, 2015*).

Missing values occur when data entries are not recorded during collection. Instead of removing such samples, appropriate imputation methods such as mean, median, or machine learning–based approaches are commonly applied (*Luengo, García & Herrera, 2012*). Proper imputation ensures that the ANN-KGA and ANN-IKGA models can learn from complete and consistent input patterns.

Outliers refer to observations that significantly deviate from the general distribution and may distort ANN learning due to their influence on weight updates (*Ayadi et al., 2017*). In this study, the Interquartile Range (IQR) method was employed. Values outside $Q_1 - 1.5\,(Q_3 - Q_1)$ and $Q_3 + 1.5\,(Q_3 - Q_1)$, were identified and treated as outliers (*Gustavsen et al., 2011*).

Data reduction

Reduction can be performed using row-based or column-based methods (*Peng & Fan, 2015*). In this study, column-based reduction—specifically feature selection—was applied to identify the most informative variables and eliminate irrelevant or redundant features, thereby improving model performance and computational efficiency; detailed information regarding the feature selection procedure and all its stages can be found in 'Feature Selection'.

Table 1 Summary of feature selection methods in the literature.

Method	Optimization objective	Approach	Strengths	Limitations
Multi-filter ranking + improved simplified swarm optimization (iSSO) + local refinement (*Lai et al., 2022*)	Metaheuristic swarm optimization (global) + local search	Single-objective (minimize prediction error) after multi-filter pre-ranking	Reduces feature count and prediction error; robust across tested sets	Heuristic tuning required; potential sensitivity to dataset characteristics
Feature-extraction (PCA and alternatives) + standard regressors (*Fan et al., 2022*)	Unsupervised linear algebraic transformations (*e.g.*, PCA)	Transform inputs to low-dim representation maximizing variance/ robustness	Improves robustness to noise and multicollinearity; reduces dimensionality	May reduce interpretability; component count choice impacts results
Model ensemble (SVR + emotional ANN) (*Hussain, Cavus & Sekeroglu, 2021*)	SVR hyperparameter tuning + ANN training (gradient-based)	Combine complementary learners to reduce model error	Stronger generalization *vs* single models; leverages complementary strengths	Not focused on feature-selection; higher complexity and compute
Fuzzy weighting + evolutionary (global/local) feature selection (*Keivanian, Chiong & Fan, 2023*)	Evolutionary algorithms with adaptive operators and fuzzy weighting	Multi-objective *via* fuzzy weighted sum (accuracy, sparsity, robustness)	Explicit multi-criteria control; balances trade-offs flexibly	Increased algorithmic complexity; fuzzy weights selection can be subjective
Parsimonious statistical regression equations (BMI + demographics) (*Xu et al., 2023*)	Classical statistical estimation (calibration + validation)	Simple, interpretable models calibrated to DXA ground truth	Highly interpretable; robust at population level; low cost	Less individualized accuracy; limited adaptability to varied populations
Regularized linear and tree-based baselines (LR, Ridge, Bayesian Ridge, trees) (*Santos, 2023*)	Convex solvers (closed form/iterative) and tree learning	Minimize regularized loss (MSE + penalty); robust baseline evaluation	Simple, reproducible baselines; strong interpretability; easy replication	May be outperformed by nonlinear/hybrid pipelines on some datasets

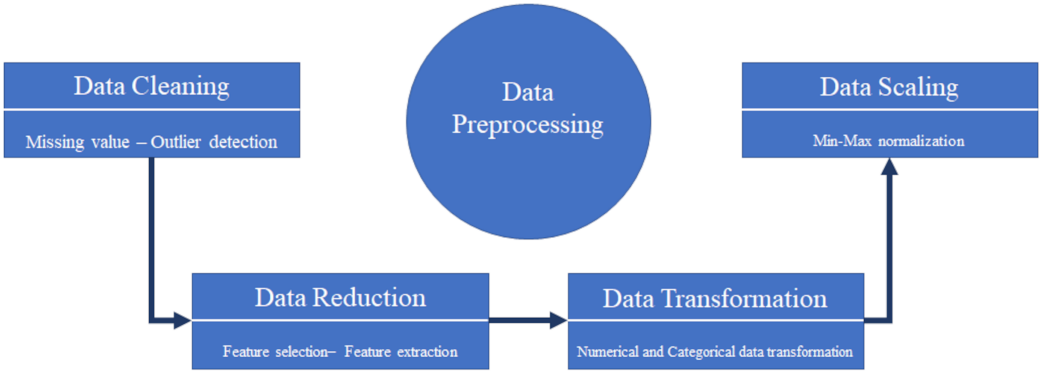

Figure 1 Typical data pre-processing tasks. Full-size DOI: 10.7717/peerj-cs.3984/fig-1

Unlike feature extraction methods that generate new variables (*Liu & Motoda, 2013*), this study focuses solely on selecting the most informative existing features through evolutionary optimization.

Data transformation

Data transformation ensures compatibility between input data and learning algorithms (*Fan et al., 2021*). Since ANN models operate directly on numerical inputs, transformation

was minimal in this study. The dataset primarily consisted of numerical variables suitable for ANN training without categorical encoding.

Data scaling

Data scaling is particularly important for ANN models because neural networks are sensitive to feature magnitude differences. Large-scale variations can negatively affect gradient-based learning. Therefore, Min–Max normalization was applied: $x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}}$ (*Jain, Nandakumar & Ross, 2005*; *Patro & Sahu, 2015*). This normalization ensures that all features contribute proportionally during ANN weight optimization. Consequently, scaling enhances convergence stability and improves the performance of both ANN-KGA and ANN-IKGA models.

Feature selection

Data reduction denotes the process of decreasing specific elements of a dataset. This may include reducing the number of data points or its dimensionality, leading to a smaller overall data size. The objectives of data reduction can differ, encompassing tasks from eliminating invalid data to producing statistical and summary information for diverse applications (*García, Luengo & Herrera, 2015*).

Data reduction is typically performed using row-based and column-based techniques. The row-based approach concentrates on decreasing the number of data samples through methods such as random and stratified sampling (*Fan et al., 2015*). Conversely, the column-based approach targets the reduction of variables within the dataset. Generally, there are three strategies for the column-based method. The first relies on domain knowledge for selecting variables. The second involves feature selection, which identifies the most significant features (*Liu & Motoda, 1998*, *2007*). The third strategy is feature extraction, which generates new features from existing ones (*Fan et al., 2021*; *Liu & Motoda, 1998*). In this study, we focus on the feature selection approach, as our primary goal is to intelligently identify important and relevant features from the body fat dataset using the ANN-IKGA.

Feature selection constitutes a vital stage in data mining, concentrating on the identification of relevant features within a dataset. It functions to distinguish pertinent features from irrelevant ones, particularly in high-dimensional datasets (*Ramchandran & Sangaiah, 2018*). Moreover, feature selection acts as a dimensionality reduction strategy, improving classification performance by eliminating unnecessary or irrelevant features (*Dey et al., 2018*). The feature selection process encompasses five essential steps, illustrated in Fig. 2, with decisions at each stage directly influencing the effectiveness of the feature selection (*Langley, 1994*). In the subsequent section, these steps will be briefly described:

- **Step 1:** Entails defining the search direction and the initial point for feature selection. The search may start with an empty set and progressively incorporate features in each iteration, referred to as forward selection. Alternatively, it can begin with a complete set and iteratively remove features, known as backward elimination. A bidirectional search

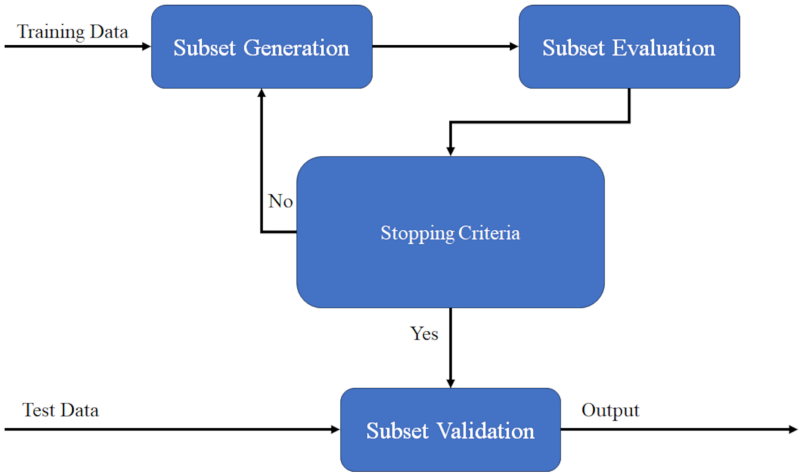

Figure 2 **Typical steps in feature selection.** Full-size DOI: 10.7717/peerj-cs.3984/fig-2

strategy adds and removes features simultaneously. Another approach involves randomly selecting features to form an initial subset (*Ang et al., 2015*).

- **Step 2:** Focuses on defining the search strategy for feature selection. An effective strategy should provide global search capabilities, rapid convergence toward the optimal solution, efficient local search, and computational efficiency (*Gheyas & Smith, 2010*). Search strategies are generally classified into three categories: exponential, sequential, and random, as examined by several studies (*Chan et al., 2010*; *Kumar & Minz, 2014*; *Muni, Pal & Das, 2006*; *Ruiz, Riquelme & Aguilar-Ruiz, 2005*).
- **Step 3:** Involves establishing the evaluation criteria to assess the effectiveness of various feature selection methods. These criteria are crucial for identifying the most suitable features. The selection of evaluation criteria depends on the specific problem and the nature of the dataset. Commonly employed evaluation approaches in feature selection include Filter (*Chandrashekar & Sahin, 2014*), Wrapper (*Kohavi & John, 1997*), Embedded (*Zheng & Casari, 2018*), and Ensemble (*Opitz & Maclin, 1999*).
- **Step 4:** Concerns establishing the stopping criteria for the feature selection process. These criteria define when the process should terminate. Appropriate stopping criteria help prevent overfitting and reduce computational complexity, facilitating a more efficient identification of the optimal feature subset. The choice of stopping criteria may be influenced by decisions made in previous steps. Common stopping criteria include having a predefined number of features, a predefined number of iterations, reaching a certain percentage of progress between two consecutive iterations, or attaining the optimal feature subset according to specific evaluation functions (*Ang et al., 2015*; *Kumar & Minz, 2014*; *Liu & Yu, 2005*):
- **Step 5:** Involves validating the final outcomes of the feature selection process. One approach to validation is to compare the selected features with prior knowledge of the dataset, if such information is available, by matching them against a known set of relevant features (*Liu & Motoda, 1998*; *Zeng et al., 2009*). However, in practical

applications, prior knowledge is often unavailable. In these cases, indirect validation methods are employed, including cross-validation, confusion matrix analysis, Jaccard similarity-based measures, and the random index. Among these, cross-validation is the most widely used technique for validation (*Bolón-Canedo, Sánchez-Maroño & Alonso-Betanzos, 2013*).

A critical aspect of feature selection, especially when embedded within an optimization framework, is the validation strategy employed to assess the selected features. To prevent optimistic bias and ensure that the feature selection process does not leak information from the test set into the model training phase, this study employs a rigorous nested cross-validation approach (*Parvandeh et al., 2020*). This methodology, detailed in 'Nested Cross-Validation Framework for Robust Evaluation', ensures that feature selection, model training, and hyperparameter tuning are performed exclusively on training data, with the final evaluation conducted on completely unseen test data.

Artificial neural network

The design of artificial neural networks (ANNs) emphasizes the management of both information flow and computational complexity among neurons. The earliest conceptual model was introduced by *McCulloch & Pitts (1943)*, which laid the groundwork for contemporary ANN architectures. With progress in neuroscience, these models developed into the commonly used structures today, comprising input, hidden, and output layers. The input layer receives preprocessed data, the hidden layer carries out the primary computations, and the output layer produces the final results (*Imran & Alsuhaibani, 2019*; *Karaman, Odabas & Turkay, 2024*).

ANNs are composed of interconnected nodes, or neurons, with weighted connections that are adjusted during training to enhance performance. Their mathematical formulation is represented in the equation below (*Lichtner-Bajjaoui, 2021*):

$$y^{(l)} = f^{l}\left(w^{(l)}h^{(l-1)} + b^{(l)}\right).$$

The above formula, represents the forward propagation step in a neural network for layer l. In this equation, $h^{(l-1)}$ denotes the output (or activation) from the preceding layer, $w^{(l)}$ is the weight matrix connecting layer $l-1$ to layer l, and $b^{(l)}$ represents the bias vector of layer l. The linear combination $w^{(l)}h^{(l-1)} + b^{(l)}$ is subsequently passed through the activation function f^{l}, which introduces non-linearity to the network. The resulting output is the activation of layer l, denoted as $y^{(l)}$.

Optimization

Optimization refers to the process of identifying the most favorable solution or outcome for a problem by either minimizing effort or maximizing benefit. It entails determining the conditions that produce the maximum or minimum value of a particular function. By applying mathematical analysis and representing the problem as an objective function, optimization seeks the best solution among all possible alternatives. This mathematical

approach is extensively employed to address problems and support decision-making across various domains (*Burke & Kendall, 2005*).

As shown in the following equation, a standard formulation of an optimization problem involves minimizing the objective function (*Sengupta, Gupta & Dutta, 2016*):

$$min f(x)\text{: } x \in R^n \quad s.t \begin{cases} g_j(x) \geq 0, & j = \{1, 2, \ldots, J\} \\ h_k(x) = 0, & k = \{1, 2, \ldots, K\} \end{cases}$$
$$x_i^{(L)} \leq x_i \leq x_i^{(u)} \quad i = \{1, 2, \ldots, n\}$$

where $f(x)$ is the optimized function, $g_j(x)$ is j inequality constraints and $h_k(x)$ represents K equality constraints.

Meta-heuristics

Historically, problem-solving methods have largely relied on intuition or meta-heuristics, often resulting in accidental discoveries. This dependence on intuition has contributed to the widespread success and adoption of meta-heuristic algorithms among researchers. These algorithms imitate natural processes, including biological, chemical, or physical phenomena, making them effective in identifying satisfactory solutions to complex problems through trial and error. The primary aim of meta-heuristics is to obtain a good solution within a reasonable timeframe (*Yang, 2010*).

Meta-heuristic algorithms are specifically designed to address complex optimization problems that classical optimization methods cannot solve effectively. They are considered highly efficient and broadly applicable, particularly for real-world problems with hybrid characteristics (*Dokeroglu, Deniz & Kiziloz, 2022*). he main advantage of meta-heuristics lies in their effectiveness and general applicability. Unlike classical optimization algorithms, meta-heuristics emphasize exploration and exploitation of the solution space, reducing the need for extensive parameter tuning and lowering the time and cost required to find improved solutions. This explains why classical optimization approaches are often unsuitable for such problems (*Ólafsson, 2006*).

The general framework for meta-heuristic algorithms can be summarized as follows (*Almufti, 2019*):

1. **Initialization:** Generate a set of candidate solutions either randomly or using a heuristic.
2. **Evaluation:** Assess the fitness of each candidate solution using an objective function.
3. **Iteration:** Repeat the following steps until a stopping criterion is satisfied:
    a. *Selection:* Choose a subset of candidate solutions based on their fitness values.
    b. *Variation:* Apply rules or operators to the selected candidates to produce new candidate solutions.
    c. *Evaluation:* Reassess the fitness of the newly generated candidate solutions using the objective function.
    d. *Update:* Replace the current set of candidate solutions with the new ones based on their fitness.

4. **Termination:** Conclude the algorithm once the stopping criterion is met.

In summary, meta-heuristic algorithms begin by initializing a set of candidate solutions, either randomly or heuristically. Each candidate is evaluated using a fitness function. The selection process identifies the best solutions, while the variation step generates new candidates by applying specific rules or operators. Finally, the update step replaces the current set of solutions with the newly generated ones based on their evaluated fitness values.

K-means genetic algorithm and improved K-means genetic algorithm

The theory of biological evolution and genetics has led to the development of genetic algorithms (GAs), which are random algorithms inspired by natural selection (*Holland, 1992*). While powerful, standard GAs can sometimes suffer from premature convergence to local optima and a lack of diversity in the population. To address these limitations, hybrid approaches have been developed that integrate other techniques to enhance the search process.

The one such powerful hybridization involves combining the global search capability of a genetic algorithm with the local search and grouping strengths of the K-means clustering algorithm (*Sinaga & Yang, 2020*). This combination leads to the K-means Genetic Algorithm (KGA) (*Krishna & Murty, 1999*). The core idea of KGA is to use K-means clustering to partition the population of solutions (chromosomes) into several distinct groups or clusters. By performing genetic operations, such as crossover, within these clusters, the algorithm can effectively explore promising regions of the search space (exploitation). A limited amount of crossover between different clusters ensures that the algorithm continues to explore new areas, maintaining population diversity (exploration) (*Anusha & Sathiaseelan, 2015*). This structured approach helps prevent the entire population from converging to a single suboptimal solution.

The procedure of KGA can be summarized as follows and also the general framework of the algorithm is shown in Fig. 3 (*Rostami & Moradi, 2014*):

1. **Initialization:** Create an initial population of individuals (chromosomes) randomly.
2. **Fitness Evaluation:** Evaluate the fitness of each individual in the population using the objective function.
3. **Clustering:** Apply the K-means algorithm to the current population to partition it into K clusters based on the similarity of the individuals.
4. **Intra-cluster Crossover and Mutation:** For each cluster, perform crossover and mutation operations among the individuals within that cluster. This intensifies the search within promising regions.
5. **Inter-cluster Crossover:** Perform a limited number of crossover operations between individuals from different clusters to encourage the exchange of information and prevent stagnation.
6. **Selection:** Create a new population by selecting the fittest individuals from the offspring and the previous generation.

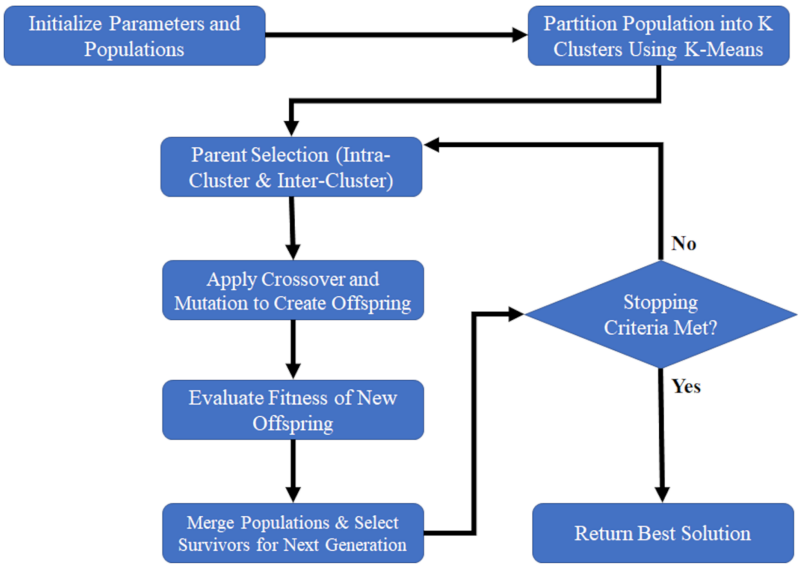

Figure 3 **General framework of the K-means genetic algorithm.**
Full-size DOI: 10.7717/peerj-cs.3984/fig-3

7. **Termination:** Check whether the stopping criteria (*e.g.*, maximum number of generations) are met. If not, go back to step 2.

Building upon the KGA, the Improved K-means Genetic Algorithm (IKGA) introduces further enhancements to better balance exploration and exploitation. A key improvement in IKGA is the dynamic adjustment of genetic operator probabilities and a more sophisticated selection mechanism (*Zhang, Ho & Lin, 2004*). For example, IKGA might use the centroid of each cluster as a representative of that region's elite solutions. The algorithm can then apply a local search mechanism around these centroids to further refine the solutions (exploitation). Additionally, IKGA can dynamically adjust the number of inter-Cluster *vs.* intra-cluster crossovers based on the population's convergence state. If the population is too diverse, it increases intra-cluster operations; if it is converging prematurely, it increases inter-cluster operations to inject diversity (*Wang & Su, 2011*). These refinements allow IKGA to adapt its search strategy throughout the evolutionary process, often leading to faster convergence and better-quality solutions compared to the standard KGA. You can see the general framework of the IKGA algorithm in Fig. 4.

Feature selection with meta-heuristic algorithms

Feature selection is a complex task; therefore, researchers increasingly rely on artificial intelligence methods to address this challenge. For years, efforts have been directed toward identifying the most effective strategies to enhance both accuracy and efficiency in this field. Among these approaches, meta-heuristic algorithms have gained prominence due to their strong effectiveness compared to other methods in improving feature selection

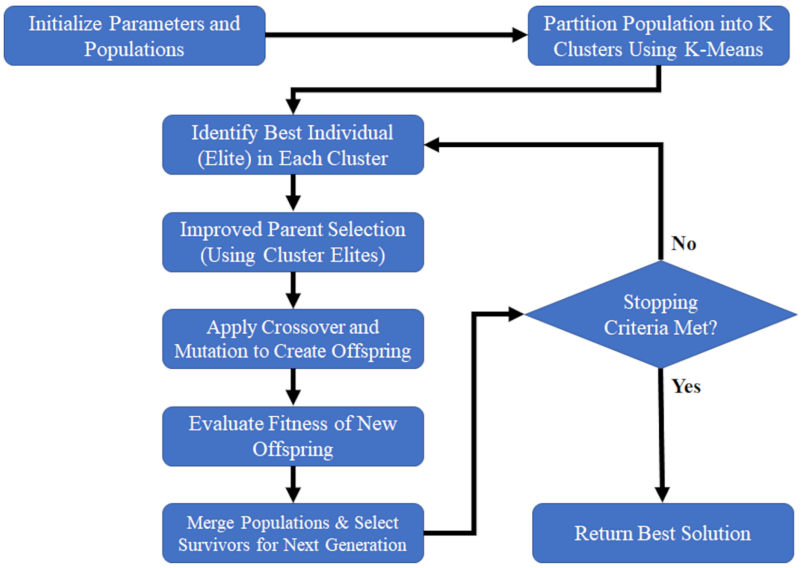

Figure 4 **General framework of the improved K-means genetic algorithm.**
Full-size DOI: 10.7717/peerj-cs.3984/fig-4

outcomes. Recent studies demonstrate the capability of these algorithms to achieve very high accuracy and significantly enhance the feature selection process (*Yusta, 2009*).

The steps for performing feature selection using meta-heuristic algorithms can be outlined as follows (*Wang et al., 2014*):

1. **Define the objective function:** The objective function evaluates the quality of the selected features, typically using a performance metric such as accuracy or R^2.
2. **Choose a meta-heuristic algorithm:** Select an algorithm based on problem-specific requirements, available computational resources, and familiarity with the method.
3. **Encoding the solution:** In meta-heuristic algorithms, solutions are usually represented as binary vectors, where each bit indicates the presence or absence of a feature.
4. **Generate an initial population:** Create an initial set of candidate solutions randomly, and evaluate each using the objective function.
5. **Iteratively search for better solutions:** Apply mutation and crossover operators to the population in order to generate new solutions. These are then evaluated using the objective function.
6. **Evaluate stopping criteria:** Terminate the algorithm once a stopping condition is satisfied, such as reaching the maximum number of iterations, achieving convergence, or exceeding computational limits.
7. **Select the best solution:** Identify the best-performing solution obtained during the search as the final feature subset.
8. **Evaluate the performance of the selected features:** Evaluate the performance of the selected features using nested cross-validation: Finally, assess the performance of the chosen features using a rigorous nested cross-validation framework. In this framework,

the feature selection process is performed exclusively within the inner loop of the cross-validation, using only the training data. The final evaluation of the selected feature subset is then conducted on the outer test fold, which has been completely isolated throughout the entire feature selection and model training process. This ensures that the reported performance metrics are unbiased and generalizable.

Proposed method

Feature selection is one of the most essential parts of data mining that many researchers work on, but a precise and appropriate solution has not been provided for this issue. For this reason, in this article, we investigated how to solve the feature selection problem using advanced hybrid meta-heuristic algorithms. ANN is a powerful machine learning algorithm for tasks like regression, while KGA and IKGA are robust evolutionary algorithms for optimization problems. To combine these algorithms for feature selection, we use the ANN as a fitness function within the KGA and IKGA frameworks. This means that the evolutionary algorithms propose candidate subsets of features, and the ANN evaluates the quality (fitness) of each subset by training a model and measuring its predictive accuracy. The goal of the KGA/IKGA is to find a feature subset that maximizes the ANN's performance while ideally minimizing the number of selected features.

Here are the general steps to hybridize ANN with KGA/IKGA:

1. **Define the Optimization Problem:** The primary objective is to maximize the prediction accuracy (or minimize the error, *e.g.*, RMSE) of the ANN model. A secondary, implicit objective is to reduce the number of features in the final subset. The problem is constrained by the total number of available features.
2. **Train the ANN as a Fitness Evaluator:** An ANN architecture is defined to take a subset of features as input and predict the body fat percentage. For each individual (feature subset) generated by the genetic algorithm, the fitness evaluation is performed using a strict 3-fold cross-validation within the outer training set. Specifically, for each candidate feature subset, the outer training data is further split into three inner folds. An ANN model is trained on two of these inner folds and evaluated on the remaining inner validation fold. This process is repeated for all three inner folds, and the fitness value for that individual is calculated as the mean R^2 score across the three inner validation folds. This ensures that the fitness evaluation is always performed on data unseen during the training of that specific ANN model, preventing any circular optimization.
3. **Integrate the ANN into KGA/IKGA:** The ANN fitness evaluation is integrated into the main loop of the KGA and IKGA algorithms. The genetic algorithm's population consists of binary vectors where each bit represents the inclusion or exclusion of a feature.
4. **Initialize the Genetic Algorithm:** Initialize the KGA/IKGA with a random population of candidate solutions (feature subsets).
5. **Evaluate Candidate Feature Subsets:** For each candidate feature subset in the population, evaluate its fitness using the trained ANN with 3-fold inner cross-validation (as described

in step 2). This inner CV evaluation ensures complete separation between the data used to train the ANN and the data used to assess its performance for fitness calculation.

6. **Evolve the Population using KGA/IKGA:** Apply the steps of the selected genetic algorithm (KGA or IKGA):

   a. Cluster the population using K-means.

   b. Apply intra-cluster and inter-cluster genetic operators (crossover and mutation) to generate a new population of feature subsets.

   c. In the IKGA variant, apply local search mechanisms or adjust operator probabilities dynamically.

7. **Repeat the Process:** Repeat steps 5–6 until a termination condition is met, such as a maximum number of generations or a desired level of convergence.

Our proposed method, as illustrated in the conceptual flow of Fig. 5, first trains a neural network with a feature subset proposed by the KGA/IKGA. It then evaluates the resulting accuracy. If the stopping condition is met, the optimal feature subset is identified. Otherwise, the process continues with the next generation of the genetic algorithm, refining the population of feature subsets until the termination criterion is satisfied and the best features are obtained.

Due to the stochastic nature of the genetic algorithm, which involves random population initialization and probabilistic crossover and mutation operations, the results of a single run may not fully represent the algorithm's true performance. Therefore, to obtain robust and reliable performance estimates, the entire nested cross-validation procedure described in the pseudocode below was repeated 15 times with different random seeds. For each independent run, the mean R^2 and mean RMSE across the five outer folds were calculated. The final reported results represent the average of these 15 independent runs, along with their standard deviations and 95% confidence intervals.

The following pseudocode outlines the combined ANN-KGA/IKGA approach for feature selection within the nested cross-validation framework, emphasizing the separation between inner-loop fitness evaluation and outer-loop final testing. The entire procedure was repeated 15 times with different random seeds to obtain robust performance estimates:
(Nested Cross-Validation Framework for ANN-KGA/IKGA):
Outer Loop (5-fold Cross-Validation):
For each outer fold (outer_train, outer_test):
Inner Loop (Feature Selection with 3-fold Cross-Validation):

1. Initialize a population P of candidate feature subsets randomly.
2. Repeat until a stopping criterion (max_it = 25) is met:

   a. For each subset s in P:

   i. Perform 3-fold cross-validation on outer_train:

      ○ For each inner fold (inner_train, inner_val):

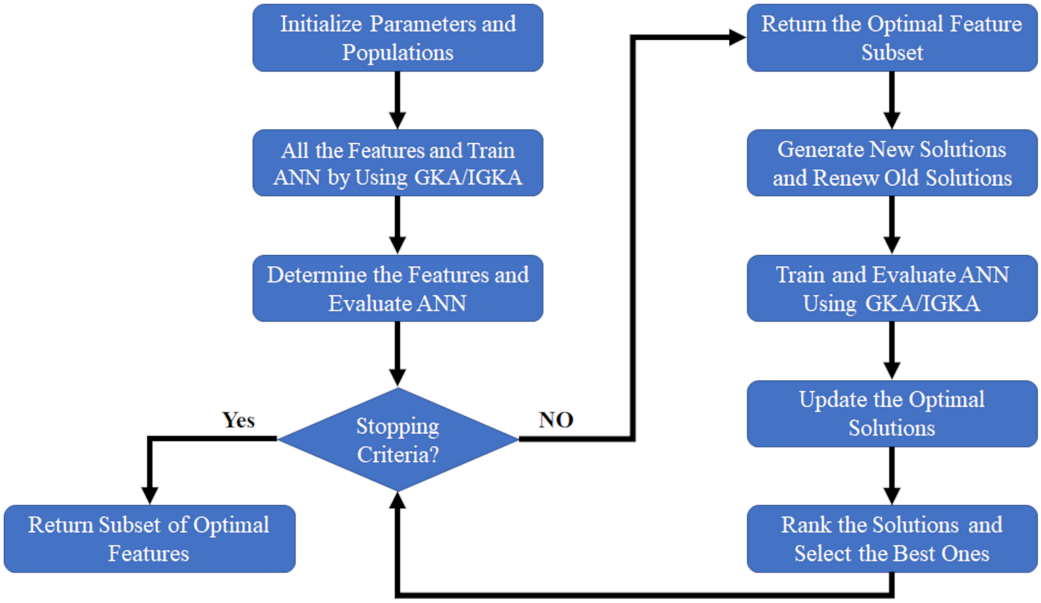

Figure 5 **General framework of the proposed algorithm.** Full-size DOI: 10.7717/peerj-cs.3984/fig-5

- Train an ANN model using only the features selected in s on inner_train.
- Evaluate the model on inner_val to obtain R^2.

ii. Calculate the fitness of s as the mean R^2 across the three inner validation folds.

b. Apply K-means to partition population P into K clusters.

c. Create an empty new population P_new.

d. Perform intra-cluster and inter-cluster crossover and mutation on P to generate offspring and add them to P_new.

e. Select the best individuals from P and P_new to form the population for the next generation.

f. Replace P with the newly formed population.

3. Select the best individual (feature subset) from the final population as the optimal feature subset for this outer fold.

Final Evaluation:

4. Train a final ANN model on the entire outer_train set using only the selected optimal features.
5. Evaluate the final model on the completely unseen outer_test set to obtain unbiased performance metrics (R^2, RMSE).

Repeat for all five outer folds and report mean performance with confidence intervals.

In this pseudocode, the KGA/IKGA is used to efficiently explore the vast search space of possible feature subsets. The clustering mechanism helps to maintain diversity and prevent premature convergence. The ANN acts as a precise evaluator to determine the actual

predictive power of each proposed subset. The final solution represents a feature subset that achieves a high predictive performance.

The time complexity of the combined algorithm depends on the number of generations G, population size N, the complexity of the K-means clustering algorithm $O(N \times K \times d \times i)$ where K is the number of clusters, d is the feature dimension, and i is the number of K-means iterations, and the complexity of the fitness evaluation (ANN training time, T_ANN). The overall time complexity can be approximated as $O(R \times F \times (G \times (N \times T\_ANN + N \times K \times d \times i)))$ where R is the number of independent runs and F is the number of outer folds. The T_ANN term is the most significant bottleneck, as it involves training a neural network for each individual in each generation. With 3-fold inner cross-validation used for fitness evaluation, each generation requires $N \times 3$ ANN trainings, making the fitness evaluation the dominant factor in the overall computational cost.

In addition to the theoretical time complexity analysis, the computational cost of the proposed method was evaluated empirically. All experiments were conducted on a system equipped with an Intel Core i7-10750H processor (2.60 GHz, 6 cores), 16 GB DDR4 RAM, and no GPU acceleration; all computations were performed on the CPU.

Due to the nested cross-validation structure and the evolutionary nature of the proposed ANN-IKGA method, the computational burden is higher compared to conventional machine learning models. For reference, the 16 baseline machine learning algorithms completed their 10-fold cross-validation hyperparameter tuning and evaluation in under 3 min in total on the same hardware. By contrast, a single run of the proposed ANN-IKGA method required approximately 18 min on average, depending on the dataset partition and stochastic initialization. Considering 15 independent runs, the total execution time was approximately 4.5 h. This additional computational cost reflects the nested evaluation of candidate feature subsets through ANN fitness evaluations across multiple generations and cross-validation folds, and represents an expected trade-off for the improved predictive performance and automatic feature selection capability of the proposed framework.

In the following section, we will discuss the advantages and disadvantages of the proposed models.

Advantages of the hybrid ANN-KGA/IKGA algorithms for feature selection:

- **Improved Solution Quality:** The combination of global search (GA) and local grouping (K-means) enhances the selection of relevant features, leading to improved solution quality.
- **Efficient Exploration and Exploitation:** KGA and IKGA maintain a better balance between exploring the entire solution space and exploiting promising regions, reducing the risk of premature convergence.
- **Enhanced Accuracy and Efficiency:** ANNs provide a robust method for evaluating fitness, enabling more accurate and efficient feature selection compared to simpler evaluation functions.
- **Versatility and Applicability: The** algorithms can be applied to various feature selection problems and are not limited to a specific domain.

Disadvantages of the hybrid ANN-KGA/IKGA algorithms for feature selection:

- **Computational Complexity:** The combination of ANNs and an iterative clustering/genetic algorithm can be computationally expensive, especially for large datasets and complex network architectures, resulting in long execution times.
- **Parameter Tuning:** The hybrid algorithms require careful tuning of multiple parameters, including population size, mutation/crossover rates, the number of clusters (K), and the neural network architecture, which can be time-consuming.
- **Interpretability Challenges:** While effective, the final feature subsets selected by the complex interplay of ANN and KGA/IKGA may lack simple interpretability.

Nested cross-validation framework for robust evaluation

To ensure a rigorous and unbiased evaluation of the proposed hybrid feature selection methods, this study employs a strict nested cross-validation (CV) strategy (*Parvandeh et al., 2020*). This approach is crucial for preventing information leakage and reducing optimistic bias, especially when feature selection is combined with model training and optimization. The nested CV (nCV) framework fully separates feature selection, hyperparameter tuning, and final model evaluation, ensuring that performance metrics reflect true generalization capability rather than data-specific artifacts. This methodology enhances the reliability and reproducibility of the experimental results.

The nested CV procedure is structured as follows:

a. **Outer Loop (Performance Evaluation):** The entire dataset (n = 252) is divided into five outer folds using stratified K-fold cross-validation to ensure balanced class distributions across folds. In each of the five iterations, one-fold (approximately 20% of the data) is held out as an outer test set. The remaining four folds (approximately 80% of the data) constitute the outer training set. All subsequent steps, including feature selection, hyperparameter tuning, and model development, are performed exclusively within this outer training set. The outer test set remains completely untouched until the final evaluation phase.

b. **Inner Loop (Feature Selection and Model Tuning):** For each outer fold, the proposed hybrid feature selection methods (ANN-KGA and ANN-IKGA) are applied only to the outer training set. To guide the evolutionary search and prevent overfitting during feature selection, an inner 3-fold cross-validation is employed:

   • **Fitness Evaluation:** For each candidate feature subset (individual) proposed by the genetic algorithm, an ANN model is trained on two of the inner folds and evaluated on the remaining inner validation fold. This process is repeated for all three inner folds.

   • **Fitness Calculation:** The fitness of a candidate feature subset is calculated as the mean R^2 score across the three inner validation folds. This ensures that the feature selection process is validated on data unseen during the training of that specific ANN model within the inner loop.

   • **Evolutionary Search:** The KGA and IKGA algorithms use this fitness value to guide crossover, mutation, and selection operations over 25 generations, with a population

size of 40. The goal is to identify the feature subset that maximizes the mean inner validation R^2.

• **Optimal Feature Subset:** At the conclusion of the inner loop evolutionary process, the best-performing feature subset (the individual with the highest mean inner validation R^2) is selected as the optimal feature set for that outer fold.

c. **Final Model Training and Evaluation:** Once the optimal feature subset is identified by the inner loop:

• **Final Model Training:** A final ANN model is trained on the entire outer training set (all four folds combined) using only the selected optimal features. The same ANN architecture and hyperparameters are used for this final training.

• **Evaluation on Unseen Data:** The trained final model is then evaluated on the completely unseen outer test set (the fold that was held out at the beginning of the outer loop). Performance metrics, including R^2 and RMSE, are calculated based solely on this evaluation.

• **Iteration:** This entire procedure (outer loop split, inner loop feature selection, final model training, and evaluation) is repeated for all five outer folds.

d. **Reporting of Results:** To account for the stochastic variability inherent in the genetic algorithm, the entire nested CV procedure (all five outer folds) was repeated independently 15 times with different random seeds. For each run, the mean R^2 and mean RMSE across the five outer folds were calculated. The final aggregated metrics reported in this study represent the average of these 15 independent runs:

• **Mean R^2 and Mean RMSE:** The average of the mean performance values from each of the 15 runs.

• **Standard Deviation (Std) across runs:** The standard deviation of the mean performance values across the 15 runs, quantifying the variability due to stochastic optimization.

• **95% Confidence Intervals (CI):** Calculated as the mean across runs ± 1.96 × (standard deviation across runs/√15), providing a measure of precision for the estimated performance.

• **Statistical Comparisons:** Paired t-tests are conducted between the R^2 values of the proposed methods (ANN-KGA/ANN-IKGA) and the baseline model using all outer-fold R^2 values from all 15 runs combined (75 paired observations: 15 runs × five folds).

• **Feature Stability:** The Jaccard similarity index is calculated across all selected feature subsets from all outer folds and all 15 runs to evaluate the consistency and robustness of the feature selection process across different random initializations.

This multiple-run approach ensures that the reported performance metrics are not dependent on a single random initialization and provide a comprehensive assessment of the algorithms' true generalization capability.

Ablation study: isolating the effect of feature selection

To directly address whether the performance improvement of ANN-IKGA stems from the feature selection mechanism itself rather than from the specific ANN architecture or preprocessing pipeline, a dedicated ablation study was conducted. This study compares two conditions under identical experimental settings:

- **Condition A (ANN-AllFeatures):** A standalone ANN model trained on all 13 original features, without any feature selection.
- **Condition B (ANN-IKGA):** The proposed hybrid model using the IKGA-selected feature subset (5–7 features per fold).

Both conditions employed the identical ANN architecture (two hidden layers with 64 neurons each, ReLU activation, Adam optimizer, max 500 iterations with early stopping), the same fold-wise preprocessing pipeline (IQR outlier capping followed by Min-Max scaling), and the identical nested cross-validation protocol (five outer folds × three inner folds). To account for stochastic variability, the entire procedure was repeated 15 times with different random seeds.

For ANN-IKGA, the fitness evaluation during the inner-loop evolutionary search (25 generations, population size 40) was performed using 3-fold cross-validation on the outer training set, with the mean R^2 across inner validation folds serving as the fitness metric. For ANN-AllFeatures, no feature selection was applied; the model was trained directly on the preprocessed feature set.

The final evaluation for both conditions was conducted on the outer test folds, which remained completely unseen during both feature selection (where applicable) and model training. Performance metrics (R^2 and RMSE) were recorded for each outer fold across all 15 runs, resulting in 75 paired observations per metric. A paired t-test was then conducted to determine whether the observed differences were statistically significant.

Model evaluation

Performance metrics, often referred to as error measures, are fundamental components in evaluation frameworks across a wide range of disciplines (*Botchkarev, 2018*). Evaluation itself is the process of gathering, analyzing, and interpreting data to assess how effectively a system operates. Through this process, the system's efficiency and overall effectiveness can be determined. As a result, evaluation becomes a key step in understanding system performance and provides valuable insights for making informed decisions and implementing improvements (*Thorpe, 1988*).

In this study, the model's performance is evaluated using four primary metrics: Mean Square Error (MSE) (*Das, Jiang & Rao, 2004*), Root Mean Square Error (RMSE) (*Willmott & Matsuura, 2005*), Mean Absolute Error (MAE) (*Chai & Draxler, 2014*), and R-Squared (R^2) (*Figueiredo Filho, Júnior & Rocha, 2011*). A detailed description of these metrics, along with their computational formulations, is presented in Table 2. All metrics are reported with their 95% confidence intervals, calculated from the five outer folds of the

Table 2 Metrics for evaluation model.

Metrics	Formula	Description
Mean Squared Error (MSE)	$\frac{1}{N}\sum_{i=1}^{N}(y_i-\hat{y})^2$	It quantifies the average of the squared differences between the predicted values and the observed actual values (*Das, Jiang & Rao, 2004*).
Root Mean Squared Error (RMSE)	$\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i-\hat{y})^2}$	It represents the square root of the mean of the squared differences between the predicted values and the actual observations (*Willmott & Matsuura, 2005*).
Mean Absolute Error (MAE)	$\frac{1}{N}\sum_{i=1}^{N}\lvert y_i-\hat{y}\rvert$	It assesses prediction accuracy by computing the mean of the absolute differences between the predicted and actual values (*Chai & Draxler, 2014*).
R-Squared (R^2)	$1-\frac{\sum_{i=1}^{N}(y_i-\hat{y})^2}{\sum_{i=1}^{N}(y_i-\bar{y})^2}$	It quantifies the proportion of variance in the dependent variable that can be explained by the independent variables in a regression model (*Figueiredo Filho, Júnior & Rocha, 2011*).

Note:
N, Number of observations; y_i, Actual value; $\hat{y}$, Predicted value; $\bar{y}$, Mean value.

nested cross-validation procedure described in 'Nested Cross-Validation Framework for Robust Evaluation'.

While discrimination metrics such as R^2, RMSE, MAE, and MSE evaluate predictive accuracy, they do not assess systematic prediction bias. Therefore, in this study, we employed calibration analysis (*Cuadros-Rodríguez et al., 2007*) to evaluate potential systematic prediction bias. Calibration examines the agreement between predicted and observed values across the prediction range. In regression models, ideal calibration occurs when predictions closely match observed outcomes without systematic deviation. The calibration slope indicates whether predictions are too extreme or too compressed, while the intercept reflects systematic upward or downward shifts. In addition, residual diagnostics (*Pagan & Hall, 1983*) were conducted in this study to examine the error structure of the model by analyzing the differences between observed and predicted values. This assessment helps identify systematic bias, heteroscedasticity, non-linearity, and deviations from distributional assumptions. Evaluating residual distribution and residual-*vs*.-fitted patterns provides insight into whether prediction errors are randomly dispersed or follow a structured pattern.

RESULTS

In this study, we aim to predict body fat percentage using the most significant features through a rigorous and unbiased evaluation framework. To ensure that all reported performance metrics are generalizable and free from optimistic bias, this study employs a strict nested cross-validation (CV) strategy for all hybrid feature selection methods (ANN-KGA and ANN-IKGA). The nested CV procedure completely isolates feature selection, model training, and hyperparameter tuning from final performance testing.

For the 16 baseline machine learning algorithms (six linear and 10 non-linear), a standard 10-fold cross-validation was applied within the training portion of each outer fold to tune their hyperparameters, with final evaluation performed on the corresponding outer test fold to maintain consistency with the nested CV framework. For the proposed hybrid methods, the nested CV procedure consists of an outer loop of five folds for performance estimation and an inner loop of three folds for feature selection and fitness evaluation. This

Table 3 Features of the body fat dataset.

Number of features	Unit of measurement
Age	Years
Weight	Pound
Height	Inches
Neck circumference	Centimeter
Chest circumference	Centimeter
Abdomen circumference	Centimeter
Hip circumference	Centimeter
Thigh circumference	Centimeter
Knee circumference	Centimeter
Ankle circumference	Centimeter
Biceps circumference	Centimeter
Forearm circumference	Centimeter
Wrist circumference	Centimeter

approach provides five independent estimates of model performance, which are then used to calculate mean performance metrics with 95% confidence intervals.

To account for the stochastic nature of the genetic algorithm and random population initialization, the entire nested cross-validation procedure was repeated 15 times with different random seeds. All results reported for the proposed hybrid methods (ANN-KGA and ANN-IKGA) represent the average of these 15 independent runs, along with their standard deviations and 95% confidence intervals, ensuring robust and reliable performance estimates.

All data preprocessing steps, including outlier capping and feature engineering, were performed exclusively within each outer training fold and applied to the corresponding outer test fold using parameters learned solely from the training data. Specifically, the preprocessing parameters (*e.g.*, means and standard deviations for normalization) were estimated using only the outer training data and subsequently applied to the outer test fold, ensuring strict separation and preventing any information leakage. Subsequently, 16 machine learning algorithms were evaluated on the dataset, followed by the implementation of our proposed hybrid feature selection methods (ANN-KGA and ANN-IKGA) within the nested CV framework. The results obtained from all algorithms are presented and compared in the following subsections.

Dataset

In this section, we first introduce the dataset employed in this study. The subsequent steps are presented sequentially to enhance the clarity and comprehension of the methodology. The dataset (*Johnson, 1996*) focuses on predicting body fat percentage based on anatomical measurements and comprises data from 252 individuals, with 13 physical characteristics recorded for each participant. The specific features included in the dataset are detailed in Table 3.

Table 4 Statistical description of body fat datasets.

	Count	Mean	Std	Min	0%	5%	50%	95%	99%	100%	Max
Body fat	252	19.15	8.37	0	0	6.055	19.2	32.6	36.621	47.5	47.5
Age	252	44.88	12.60	22	22	25	43	67	72	81	81
Weight	252	81.16	13.33	53.75	53.75	61.86	80.06	102.35	111.46	164.72	164.72
Height	252	178.18	9.30	74.93	74.93	167.35	177.80	189.23	193.04	197.49	197.49
Neck	252	37.99	2.43	31.1	31.1	34.255	38	41.845	42.996	51.2	51.2
Chest	252	100.82	8.43	79.3	79.3	89.02	99.65	116.34	120.73	136.2	136.2
Abdomen	252	92.56	10.78	69.4	69.4	76.875	90.95	110.76	120.01	148.1	148.1
Hip	252	99.90	7.16	85	85	89.155	99.3	112.13	115.79	147.7	147.7
Thigh	252	59.41	5.25	47.2	47.2	51.155	59	68.545	72.696	87.3	87.3
Knee	252	38.59	2.41	33	33	34.8	38.5	42.645	44.592	49.1	49.1
Ankle	252	23.10	1.69	19.1	19.1	21	22.8	25.445	28.274	33.9	33.9
Biceps	252	32.27	3.02	24.8	24.8	27.61	32.05	37.2	38.5	45	45
Forearm	252	28.66	2.02	21	21	25.7	28.7	31.745	33.394	34.9	34.9
Wrist	252	18.23	0.93	15.8	15.8	16.8	18.3	19.8	20.645	21.4	21.4

The descriptive statistics (mean, standard deviation, minimum, and maximum) for all 13 original anthropometric features and the target variable (body fat percentage) are presented in Table 4. These statistics provide essential context for understanding the distribution and range of the data used in this study. The target variable, body fat percentage, has a mean of 19.15% and a standard deviation of 8.37%. This variance ($\sigma^2 = 70.06$) provides essential context for interpreting the prediction performance metrics reported in 'Body Fat Prediction with Proposed Methods (ANN-KGA & ANN-IKGA)'.

The target population of this study consists of adult individuals for whom standard anthropometric measurements are available. The model is intended for use in general adult populations rather than specific clinical subgroups. Therefore, the findings should be interpreted within the context of non-invasive body composition assessment in routine healthcare and public health settings. Extrapolation to pediatric populations or patients with severe metabolic or endocrine disorders requires further validation.

Body fat prediction with machine learning algorithms

This part of the study applies 16 machine learning algorithms to the body fat dataset, including six linear and 10 nonlinear techniques. In this experimental setting, all models were trained and evaluated using the full feature set, without the application of any feature selection method, in order to provide a consistent baseline for comparison.

Initially, we employed a set of linear models—namely Multiple Linear Regression (MLR) (*Draper & Smith, 1998*), Partial Least Squares Regression (PLS) (*Wold, Sjöström & Eriksson, 2001*), Ridge Regression (*Hoerl & Kennard, 1970*), Lasso Regression (*Tibshirani, 1996*), Elastic-Net Regression (*Zou & Hastie, 2005*), and Support Vector Regression (SVR) (*Drucker et al., 1996*)—to estimate body fat. The outcomes of these linear approaches are presented in Table 5.

Table 5 Results obtained from linear models for predicting body fat dataset.

	MSE train	RMSE train	MSE test	RMSE test	MAE train	MAE test	R^2 train	R^2 test
MLR	17.934	4.325	19.150	4.376	3.504	3.470	0.724	0.763
PLS	17.934	4.235	19.212	4.383	3.506	3.473	0.724	0.762
Ridge	18.014	4.244	19.746	4.443	3.496	3.517	0.723	0.755
Lasso	18.585	4.311	21.119	4.596	3.601	3.573	0.714	0.738
E-Net	18.163	4.262	19.611	4.428	3.544	3.498	0.721	0.757
SVR	19.004	4.359	19.889	4.460	3.464	3.553	0.708	0.754

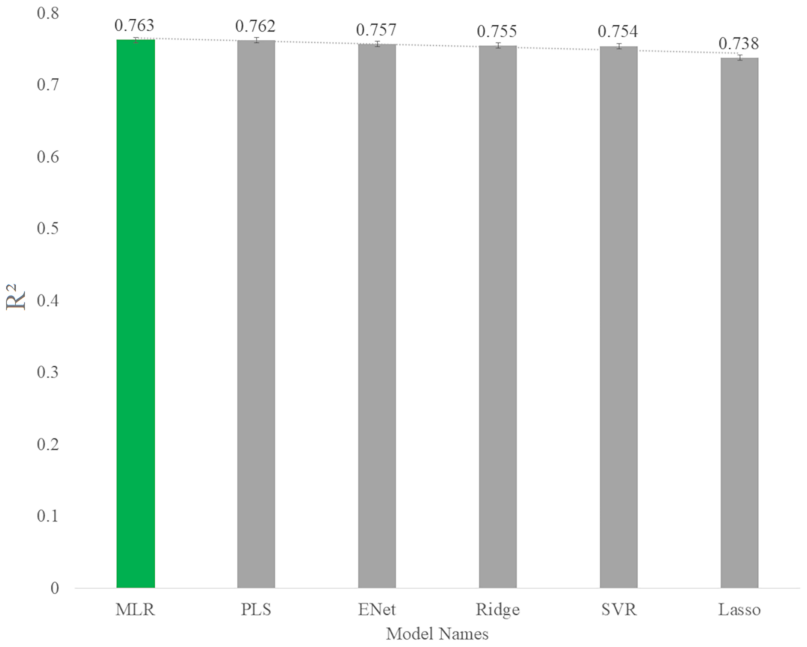

Figure 6 Performance of linear models in body fat estimation.
Full-size DOI: 10.7717/peerj-cs.3984/fig-6

As illustrated in Fig. 6, within the group of linear algorithms, multiple linear regression achieves the highest prediction performance on the body fat dataset, with an R^2 of 0.763.

In the subsequent step, we employ non-linear algorithms, including K-Nearest Neighbors (KNN) (*Fix & Hodges, 1989*), Non-Linear Support Vector Regression (N-SVR) (*Cortes & Vapnik, 1995*), Artificial Neural Network (ANN) (*McCulloch & Pitts, 1943*), Classification and Regression Trees (CART) (*Breiman et al., 2017*), Bagged Trees (*Breiman, 1996*), Random Forests (RF) (*Breiman, 1996*), Gradient Boosting Machines (GBM) (*Friedman, 2001*), Extreme Gradient Boosting (XGBoost) (*Chen et al., 2015*), LightGBM (*Ke et al., 2017*), and Category Boosting (CatBoost) (*Prokhorenkova et al., 2018*) for body fat prediction. The results derived from these non-linear models are presented in Table 6.

Table 6 Results obtained from non-linear models for predicting body fat dataset.

	MSE train	RMSE train	MSE test	RMSE test	MAE train	MAE test	R^2 train	R^2 test
KNN	25.268	5.027	24.950	4.995	4.124	4.179	0.611	0.691
N-SVR	21.646	4.653	18.562	4.297	3.744	3.592	0.667	0.771
ANN	24.520	4.952	27.984	5.290	4.109	4.097	0.623	0.653
CART	23.608	4.859	27.610	5.255	3.808	4.425	0.637	0.658
BGTrees	4.023	2.006	22.395	4.732	1.527	4.010	0.938	0.723
RF	4.113	2.028	22.090	4.700	1.639	3.964	0.937	0.726
GBM	6.476	2.543	23.680	4.866	2.112	4.097	0.901	0.707
XGBoost	7.557	2.749	24.121	4.911	2.188	4.020	0.884	0.701
LGBM	11.856	3.443	22.508	4.744	2.751	3.997	0.818	0.721
CatBoost	3.641	1.908	22.918	4.787	1.559	4.084	0.944	0.716

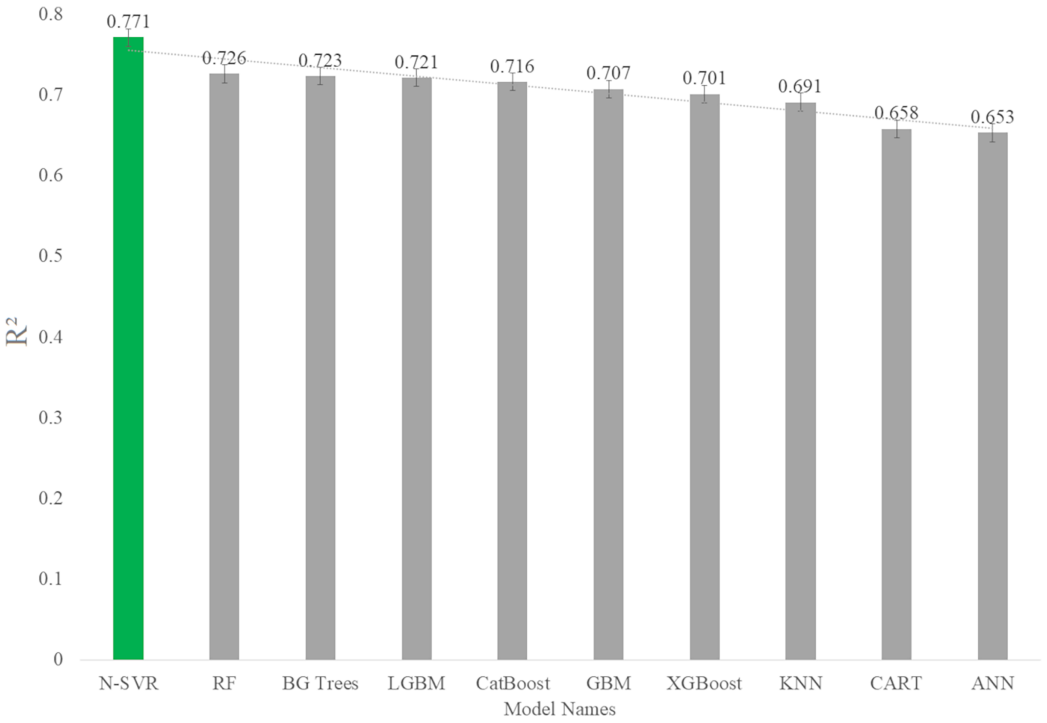

Figure 7 Performance of non-linear models in body fat estimation.
Full-size DOI: 10.7717/peerj-cs.3984/fig-7

As illustrated in Fig. 7, among the non-linear algorithms, Non-Linear Support Vector Regression achieves the highest prediction performance on the body fat dataset, with an R^2 of 0.77.

All hyperparameters and configuration settings obtained for the sixteen machine learning algorithms evaluated in this study—including model-specific tuning parameters, optimization settings, and structural configurations—are comprehensively detailed in Table 7. These parameters represent the best-performing hyperparameter combinations

Table 7 Hyperparameter settings and configuration details used for the sixteen different machine learning algorithms.

Method	Hyperparameters
MLP	activation: relu, alpha: 0.0001, hidden_layer_sizes: (100,), learning_rate: constant, learning_rate_init: 0.001, max_iter: 200, solver: adam
PLS	max_iter: 500, n_components: 2
Ridge	alpha: 60.68
Lasso	alpha: 0.22
ElasticNet	alpha: 0.28
GBM	learning_rate: 0.01, max_depth: 5, n_estimators: 500, subsample: 0.5
Adaboost	learning_rate: 0.1, loss: exponential, n_estimators: 500
RF	max_depth: 5, max_features: 8, min_samples_split: 8, n_estimators: 100
BGTress	max_features: 12, n_estimators: 9
ANN	alpha: 0.2, hidden_layer_sizes: (100, 100, 50)
XGBoost	colsample_bytree: 0.5, learning_rate: 0.1, max_depth: 5, n_estimators: 200
CatBoost	depth: 5, iterations: 200, learning_rate: 0.1
KNN	n_neighbors: 6
CART	max_leaf_nodes: 4, min_samples_split: 12
SVR	C: 1.90
LGBM	colsample_bytree: 0.75, learning_rate: 0.01, max_depth: 5, n_estimators: 500

Table 8 Hyperparameter settings and architectural configurations used for the proposed models.

Method	Hidden layer size	Activation function	Solver	ANN-iteration	Early stopping	KGA & IKGA iteration	Population	K (cluster number)
ANN-KGA & ANN-IKGA	(64, 32, 16)	Relu	Adam	300	True	25	40	3

identified during the model selection process and outline the computational characteristics that governed the behavior of each algorithm. Presenting these configurations ensures methodological transparency and provides a clear basis for reproducing the experimental results.

Body fat prediction with proposed methods (ANN-KGA & ANN-IKGA)

All hyperparameters and architectural configurations used in both the standalone ANN models and the hybrid ANN–KGA and ANN–IKGA systems—including the number of layers, activation functions, iteration counts, and all additional optimization settings—are comprehensively detailed in Table 8. These parameters define the structural and computational characteristics of the proposed framework and ensure transparency in the model development process.

The proposed hybrid methods (ANN-KGA and ANN-IKGA) were evaluated using the strict nested cross-validation framework. For each of the five outer folds, the inner loop evolutionary search (25 generations with a population size of 40) identified an optimal feature subset based on the mean inner validation R^2. A final ANN model was then trained on the entire outer training set using only these selected features and evaluated on the

Table 9 Performance comparison of ANN-KGA and ANN-IKGA models based on nested cross-validation results obtained from 15 independent runs.

	Mean R2	Std R^2	Mean RMSE	Feature stability	Paired t-test
ANN-KGA	0.7522 (95% CI [0.6963–0.8081])	0.0570	3.9651 (95% CI [3.5453–4.3849])	0.5742	0.0453
ANN-IKGA	0.7858 (95% CI [0.7495–0.8222])	0.0420	3.7182 (95% CI [3.1777–4.2586])	0.6821	0.0397

completely unseen outer test set. This process was repeated for all five outer folds, yielding five independent performance estimates for each method.

To account for the stochastic nature of the genetic algorithm and the random population initialization, the entire nested cross-validation procedure (five outer folds) was repeated 15 independent times with different random seeds. For each independent run, the mean R^2 and mean RMSE across the five outer folds were calculated. The final performance metrics reported in this section represent the average of these 15 independent runs, along with their standard deviations and 95% confidence intervals. This multiple-run approach ensures that the reported results are robust and not dependent on a single random initialization of the evolutionary search.

The performance comparison of the ANN-KGA and ANN-IKGA models averaged over 15 independent runs is presented in Table 9. ANN-KGA achieved a mean outer-fold R^2 of 0.7522 (95% CI [0.6963–0.8081]) and a mean RMSE of 3.9651 (95% CI [3.5453–4.3849]). In contrast, the proposed ANN-IKGA demonstrated stronger predictive performance, yielding a higher mean outer-fold R^2 of 0.7858 (95% CI [0.7495–0.8222]) and a lower mean RMSE of 3.7182 (95% CI [3.1777–4.2586]) across the 15 independent runs. Furthermore, paired t-tests comparing all outer-fold R^2 values from all 15 runs (75 paired observations: 15 runs × five folds) for both hybrid models with the baseline model (using all features) revealed statistically significant improvements for ANN-KGA ($p = 0.0453$) and ANN-IKGA ($p = 0.0397$). These findings indicate that the observed performance enhancements are statistically meaningful and reflect genuine improvements in predictive accuracy rather than random variation or artifacts of a particular random initialization.

As illustrated in Table 9, the R^2 values achieved by the proposed hybrid methods, particularly ANN-IKGA, are higher than to those obtained by the 16 baseline machine learning algorithms. The ANN-IKGA model achieved the highest overall predictive performance, with a mean R^2 of 0.7858 across the five outer folds across the 15 independent runs. This performance was attained while utilizing a substantially reduced feature set. Across the five outer folds and 15 independent runs, the number of selected features ranged from five to eight for ANN-KGA and from five to seven for ANN-IKGA, demonstrating significant dimensionality reduction from the original 13 candidate features. The consistency of feature selection across folds, measured by the Jaccard similarity index, was 0.5742 for ANN-KGA and 0.6821 for ANN-IKGA, indicating the robustness and stability of the selected feature subsets.

To verify the consistency between the reported R^2 and RMSE values, we can use the outcome variance reported in Table 4. Given the standard deviation of body fat percentage ($\sigma = 8.37\%$, $\sigma^2 = 70.06$), the relationship $R^2 = 1 - (RMSE^2/\sigma^2)$ should approximately hold. For ANN-IKGA, RMSE = 3.7182, so $RMSE^2 = 13.83$. Thus, the expected R^2 based on RMSE would be $1 - (13.83/70.06) = 1 - 0.197 = 0.803$. Our empirical R^2 of 0.7858 is slightly lower than this theoretical expectation, which is expected due to sampling variability across different folds and runs. This minor discrepancy confirms that the reported metrics are mutually consistent and that the model performs as expected given the inherent variance in the dataset.

This rigorous multiple-run nested CV procedure, with fitness evaluation based on inner validation folds completely separate from the outer test sets and results averaged over 15 independent runs, guarantees that the reported performance metrics (mean $R^2 = 0.7858$, std = 0.0420, 95% CI [0.7495–0.8222] for ANN-IKGA) are unbiased, robust to stochastic variability, and reflect true generalization capability, free from any circular optimization or data leakage.

In addition to discrimination metrics, the calibration performance of the ANN-IKGA model was evaluated on external test folds. The calibration slope was 0.9278 and the calibration intercept was 1.8238, indicating that the model did not exhibit systematic prediction bias and demonstrated relatively robust calibration against overfitting. Furthermore, the mean prediction deviation was 0.4783, showing that the model overestimated body fat percentage by less than half a percentage point on average. The magnitude of this deviation is small relative to the standard deviation of the observed outcomes ($\sigma = 8.37$), suggesting an acceptable level of practical calibration.

The calibration plot of the ANN-IKGA model (Fig. 8) illustrates the relationship between predicted and observed values. The solid regression line represents this relationship and demonstrates an overall good fit, with only minor deviations reflected in the slope and intercept values. Additionally, residual diagnostics were conducted to further evaluate model behavior (Fig. 9). The Shapiro-Wilk test produced a *p*-value of 0.000159, indicating a statistically significant deviation from normality. However, visual inspection of the residual distribution and residual-*vs*.-fitted plots did not reveal strong heteroskedastic patterns or pronounced systematic bias. Residual variance remained relatively stable across prediction levels, suggesting an acceptable error structure for practical modeling purposes.

Ablation study results: ANN-AllFeatures *vs*. ANN-IKGA

To isolate the effect of feature selection from the influence of the underlying predictive model, a direct ablation comparison was performed between a standalone ANN trained on all 13 features (ANN-AllFeatures) and the proposed ANN-IKGA model. Both models shared the identical ANN architecture, preprocessing pipeline, nested cross-validation framework, and 15 independent runs with different random seeds.

The results of this ablation study are presented in Fig. 10. ANN-AllFeatures achieved a mean outer-fold R^2 of 0.7145 (95% CI [0.6943–0.7347]) with a mean RMSE of 4.4831 (95% CI [4.1561–4.8101]). In comparison, ANN-IKGA achieved a substantially higher mean R^2

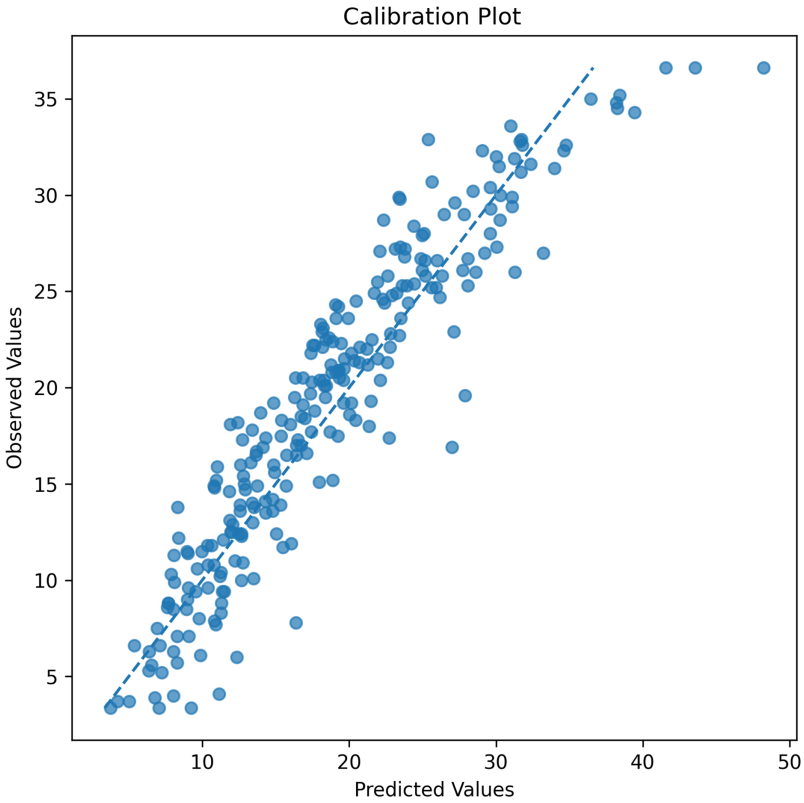

Figure 8 **Calibration plot of the ANN-IKGA model.** Full-size DOI: 10.7717/peerj-cs.3984/fig-8

of 0.7858 (95% CI [0.7495–0.8222]) and a lower mean RMSE of 3.7182 (95% CI [3.1777–4.2586]).

A paired t-test conducted on the 75 paired R^2 observations (15 runs × five outer folds) yielded a p-value of 0.0397, indicating that the observed performance difference is statistically significant at the $\alpha = 0.05$ level. This finding directly demonstrates that the feature selection mechanism inherent to the IKGA framework is the primary driver of the predictive improvement, rather than the ANN architecture or preprocessing steps alone.

DISCUSSION

In the preceding sections, a detailed analysis of machine learning-based algorithms, including the proposed hybrid model ANN-IKGA, was presented. This section now turns to a comparison of results reported by other authors using the same dataset. The growing number of studies on body fat percentage estimation reflects an increasing academic interest and expert involvement in this field (*Aristizabal, Estrada-Restrepo & Giraldo García, 2018*; *Chiong et al., 2021*; *Hastuti et al., 2018*; *Henry et al., 2018*; *Lai et al., 2022*; *Salamunes, Stadnik & Neves, 2018*; *Shao, 2014*; *Uçar et al., 2021*). Since the calculation of body fat percentage is both complex and costly, the demand for efficient and cost-effective methods continues to rise. When proven practical, such methods can contribute significantly to economic growth by enabling accurate measurements without requiring specialized equipment, thereby saving both time and cost.

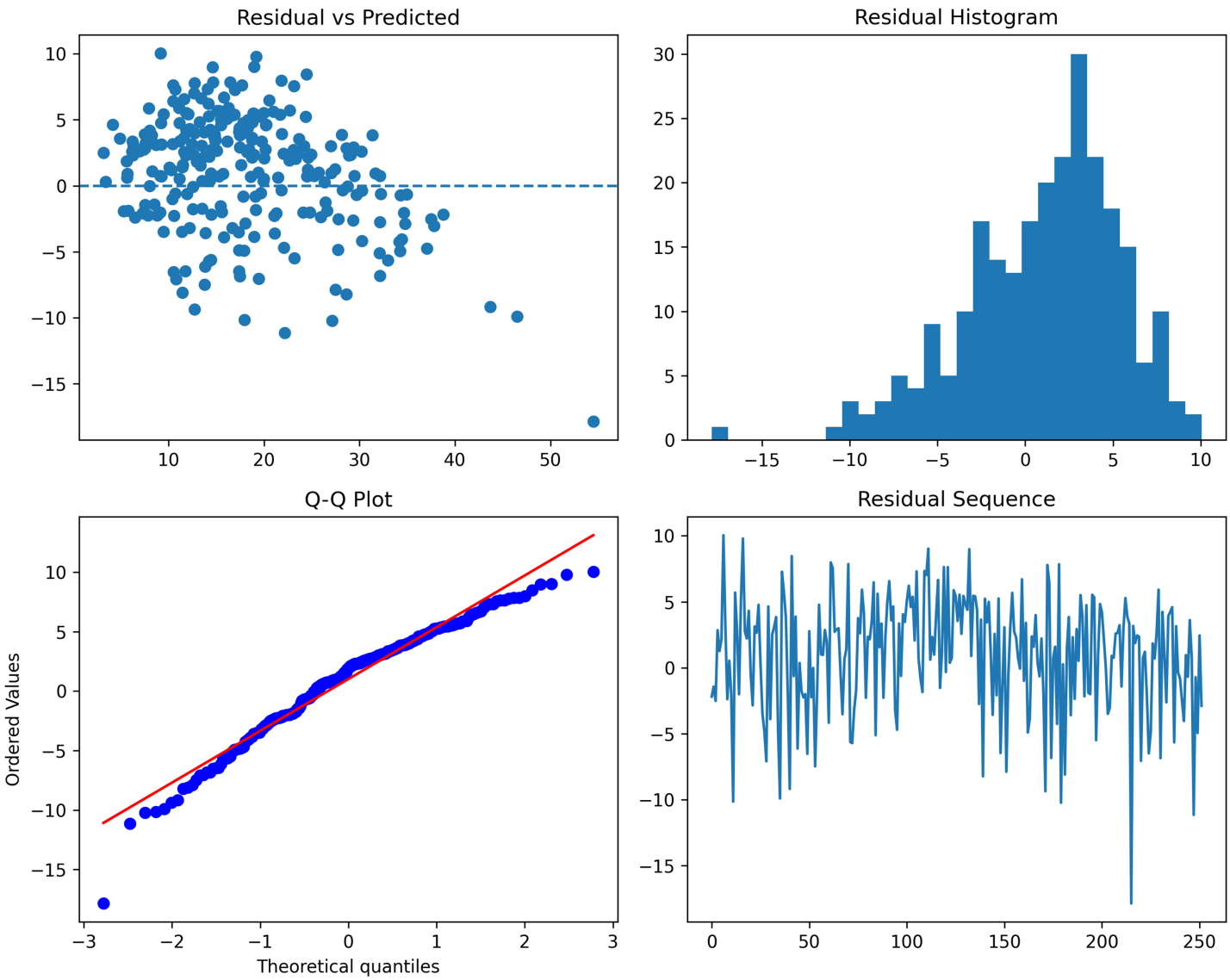

Figure 9 **Residual diagnostic plots for the ANN-IKGA model.** Full-size DOI: 10.7717/peerj-cs.3984/fig-9

It is important to emphasize that these comparisons should be interpreted with caution, as prior studies may differ in terms of preprocessing procedures, feature engineering strategies, and validation methodologies. In particular, many existing works rely on single train-test splits, whereas the present study employs a strict nested cross-validation framework with multiple independent runs, providing a more robust and unbiased estimate of generalization performance.

In this study, the proposed hybrid method (ANN-IKGA) achieved a mean RMSE of 3.7182 (95% CI [3.1777–4.2586]) and a mean R^2 of 0.7858 (95% CI [0.7495–0.8222]) under a strict nested cross-validation framework, with results averaged over 15 independent runs to account for the stochastic variability inherent in the genetic algorithm. These results represent the unbiased, generalizable performance of the model on unseen data. A key

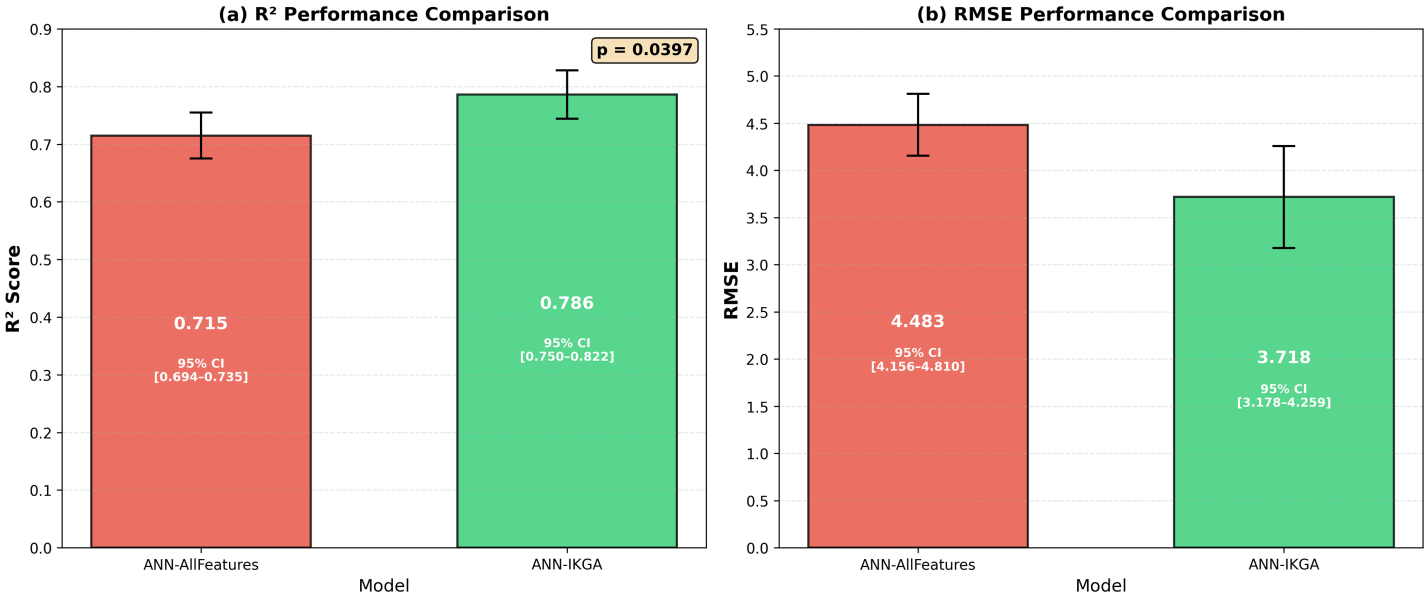

Figure 10 **Ablation study: ANN-AllFeatures *vs.* ANN-IKGA.** Full-size DOI: 10.7717/peerj-cs.3984/fig-10

distinction of this result is that such performance was obtained while substantially reducing the feature dimensionality, with the optimal feature subsets selected automatically by the method across different outer folds. This highlights the novelty of our approach, as it not only determines the most informative features automatically through an evolutionary search but also delivers robust and accurate estimation of body fat, setting it apart from existing studies in the literature that often rely on single train-test splits and may report optimistically biased results.

The hybrid method proposed in this study achieved an mean RMSE of 3.7182, whereas *Shao (2014)* reported an RMSE of 4.6384 for a hybrid approach utilizing all available features. It is important to emphasize that Shao's method incorporated the complete feature set, while our approach achieves a lower RMSE while employing a selective feature subset, reducing computational burden through dimensionality reduction.

Uçar et al. (2021) proposed a hybrid machine learning framework that combined MLFFNN, DT, and SVMs to predict body fat, assessing the predictive contribution of each feature. Using the same dataset as in our study, their most successful model reported an RMSE of 4.453 with six selected features. By comparison, our approach attained a lower mean RMSE of 3.7182 under the present nested cross-validation framework, though direct comparisons should be interpreted cautiously given potential differences in preprocessing and validation protocols.

Chiong et al. (2021) proposed a method called IRE-SVM for body fat estimation. By applying feature selection within their approach, they reported an RMSE of 4.3391, which is higher than the mean RMSE of 3.7182 achieved in our study under the present nested cross-validation framework.

In this investigation, *Lai et al. (2022)* proposed a hybrid method (VMFET-iSSO) that incorporates an enhanced version of the simplified swarm optimization algorithm.

Their findings reported an RMSE value of 4.1182, which is higher than the mean RMSE of 3.7182 achieved in our study. This further highlights the importance of our approach in accurately predicting body fat and contributing to the understanding and management of obesity.

From a clinical perspective, the achieved mean RMSE of 3.7182 corresponds to an average absolute deviation of approximately ±3% in body fat percentage estimation. In practical terms, such an error margin is generally considered acceptable for population-level screening, routine monitoring, and epidemiological research, where small individual deviations do not substantially alter overall risk stratification. However, for individual diagnostic decisions requiring high precision, this level of error may not be sufficient to replace gold-standard techniques. Therefore, the proposed model should be interpreted as a supportive decision tool for screening and health monitoring rather than a standalone diagnostic instrument. The improvement in R^2 and reduction in RMSE compared to previous approaches gains practical significance when viewed within this context of cost-effectiveness, accessibility, and large-scale applicability.

CONCLUSIONS AND FUTURE WORKS

In this study, we proposed two hybrid feature selection methods by integrating an ANN with the KGA and its improved variant (IKGA). To establish a baseline under a consistent evaluation framework, body fat percentage was first predicted using 16 machine learning algorithms within a nested cross-validation structure, where the nonlinear support vector machine achieved the highest mean R^2 of 0.77. In contrast, our proposed ANN-IKGA method, evaluated under the same strict nested cross-validation protocol and averaged over 15 independent runs, achieved a higher mean R^2 of 0.7858 (95% CI [0.7495–0.8222]) and a mean RMSE of 3.7182 (95% CI [3.1777–4.2586]) while utilizing a substantially reduced feature set automatically selected through an evolutionary search strategy. This underscores the predictive strength and robustness of our approach as well as its efficiency in intelligent feature selection. The nested cross-validation framework ensures unbiased performance estimation and provides confidence intervals, enhancing transparency and reproducibility. Comparative results in Table 9 and Fig. 11 demonstrate the advantage of the proposed hybrid method in both predictive accuracy and dimensionality reduction, distinguishing it from prior studies that do not employ similarly rigorous validation protocols.

A dedicated ablation study, directly comparing the same ANN architecture trained on all features *vs.* ANN-IKGA under identical nested cross-validation conditions, confirmed that the observed performance improvement is attributable to the feature selection mechanism itself. This finding isolates the effect of feature selection from other confounding factors and provides robust evidence for the effectiveness of the proposed hybrid framework.

Our findings reveal that utilizing all available features in a dataset does not inherently guarantee the highest R^2, underscoring the importance of intelligent feature selection methods. The proposed hybrid approach effectively identifies the most informative features, thereby constructing more robust and predictive models. Moreover, the integration of K-means clustering within the genetic algorithm framework demonstrated

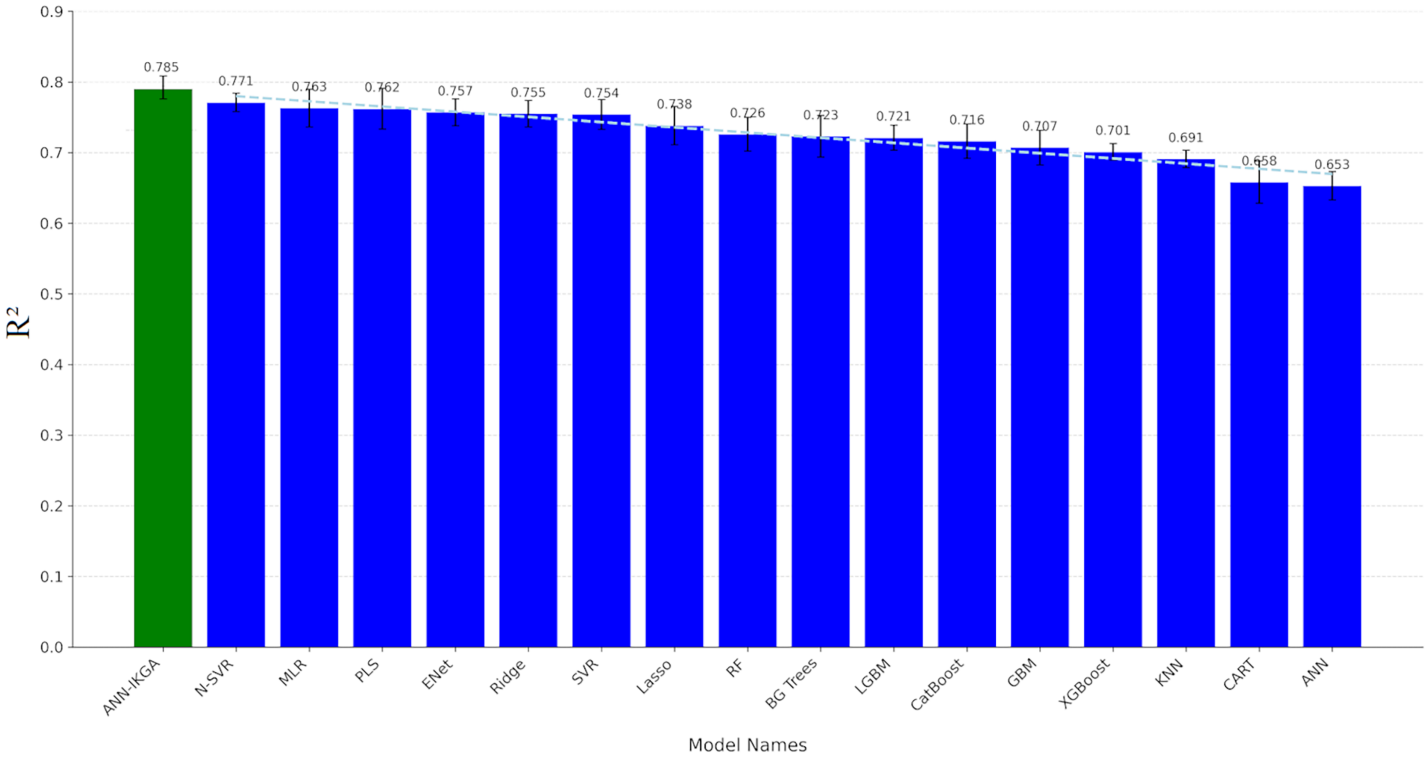

Figure 11 **Comparison of linear, non-linear, and proposed methods.** Full-size DOI: 10.7717/peerj-cs.3984/fig-11

significant benefits, mitigating premature convergence and enabling a more comprehensive exploration of the solution space. Further refinements introduced in the IKGA resulted in even greater predictive performance, highlighting the effectiveness of adaptive and advanced meta-heuristic strategies.

The primary aim of this study is to enhance feature selection by introducing our proposed hybrid methods. For future work, a promising direction would be to integrate these methods with alternative machine learning models beyond ANNs, such as Gradient Boosting Machines or Support Vector Machines, which can serve as fitness evaluators. Moreover, applying the proposed approaches to diverse high-dimensional datasets would allow us to assess their robustness, scalability, and effectiveness across multiple domains. Future research should prioritize external validation using independent datasets from diverse populations to further establish the model's generalizability and clinical robustness. Such comparative investigations across different datasets and populations could offer deeper insights into the adaptability and overall performance of the hybrid framework.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

The authors received no funding for this work.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- Tohid Yousefi conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Özlem Varlıklar conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The data and code are available in the Supplemental Files.

The data is available at Kaggle: https://www.kaggle.com/datasets/fedesoriano/body-fat-prediction-dataset.

The data is available at: https://www.mathworks.com/help/deeplearning/ug/body-fat-estimation.html.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj-cs.3984#supplemental-information.

REFERENCES

Almufti SM. 2019. Historical survey on metaheuristics algorithms. *International Journal of Scientific World* **7(1)**:1–12 DOI 10.14419/ijsw.v7i1.29497.

Ang JC, Mirzal A, Haron H, Hamed HNA. 2015. Supervised, unsupervised, and semi-supervised feature selection: a review on gene selection. *IEEE/ACM Transactions on Computational Biology Bioinformatics* **13(5)**:971–989 DOI 10.1109/tcbb.2015.2478454.

Anusha M, Sathiaseelan J. 2015. Feature selection using k-means genetic algorithm for multi-objective optimization. *Procedia Computer Science* **57**:1074–1080 DOI 10.1016/j.procs.2015.07.387.

Aristizabal JC, Estrada-Restrepo A, Giraldo García A. 2018. Development and validation of anthropometric equations to estimate body composition in adult women. *Colombia Médica* **49(2)**:154–159 DOI 10.25100/cm.v49i2.3643.

Ayadi A, Ghorbel O, Obeid AM, Abid M. 2017. Outlier detection approaches for wireless sensor networks: a survey. *Computer Networks* **129(4)**:319–333 DOI 10.1016/j.comnet.2017.10.007.

Bolón-Canedo V, Sánchez-Maroño N, Alonso-Betanzos A. 2013. A review of feature selection methods on synthetic data. *Knowledge Information Systems* **34(3)**:483–519 DOI 10.1007/s10115-012-0487-8.

Botchkarev A. 2018. Performance metrics (error measures) in machine learning regression, forecasting and prognostics: properties and typology. ArXiv DOI 10.48550/arXiv.1809.03006.

Breiman L. 1996. Bagging predictors. *Machine Learning* **24(2)**:123–140 DOI 10.1023/a:1018054314350.

Breiman L, Friedman JH, Olshen RA, Stone CJ. 2017. *Classification and regression trees*. Oxfordshire, UK: Routledge.

Burke EK, Kendall G. 2005. *Search methodologies*. Cham: Springer.

Chai T, Draxler RR. 2014. Root mean square error (RMSE) or mean absolute error (MAE)?–arguments against avoiding RMSE in the literature. *Geoscientific Model Development* **7(3)**:1247–1250 DOI 10.5194/gmd-7-1247-2014.

Chan Y-H, Ng WW, Yeung DS, Chan PP. 2010. Empirical comparison of forward and backward search strategies in L-GEM based feature selection with RBFNN. In: *2010 International Conference on Machine Learning and Cybernetics*. Piscataway: IEEE, 1524–1527.

Chandrashekar G, Sahin F. 2014. A survey on feature selection methods. *Computers & Electrical Engineering* **40(1)**:16–28 DOI 10.1016/j.compeleceng.2013.11.024.

Chen T, He T, Benesty M, Khotilovich V, Tang Y, Cho H, Chen K. 2015. XGBoost: extreme gradient boosting. R package version 04-2 1: 1–4.

Chinneck J. 2006. *Practical optimization: a gentle introduction*. Ottawa, Ontario: Carleton University.

Chiong R, Fan Z, Hu Z, Chiong F. 2021. Using an improved relative error support vector machine for body fat prediction. *Computer Methods and Programs in Biomedicine* **198(3)**:105749 DOI 10.1016/j.cmpb.2020.105749.

Coello CAC. 2007. *Evolutionary algorithms for solving multi-objective problems*. Cham: Springer.

Cortes C, Vapnik V. 1995. Support-vector networks. *Machine Learning* **20(3)**:273–297 DOI 10.1023/a:1022627411411.

Cuadros-Rodríguez L, Bagur-González MG, Sánchez-Vinas M, González-Casado A, Gómez-Sáez AM. 2007. Principles of analytical calibration/quantification for the separation sciences. *Journal of Chromatography A* **1158**:33–46 DOI 10.1016/j.chroma.2007.03.030.

Das K, Jiang J, Rao JNK. 2004. Mean squared error of empirical predictor. *The Annals of Statistics* **32(2)**:818–840 DOI 10.1214/009053604000000201.

Dey N, Borra S, Ashour AS, Shi F. 2018. *Machine learning in bio-signal analysis and diagnostic imaging*. Amsterdam, Netherlands: Academic Press.

Dokeroglu T, Deniz A, Kiziloz HE. 2022. A comprehensive survey on recent metaheuristics for feature selection. *Neurocomputing* **494(13)**:269–296 DOI 10.1016/j.neucom.2022.04.083.

Draper NR, Smith H. 1998. *Applied regression analysis*. Hoboken, New Jersey: John Wiley & Sons.

Drucker H, Burges CJ, Kaufman L, Smola A, Vapnik V. 1996. Support vector regression machines. In: *Advances in Neural Information Processing Systems 9*.

Ellis KJ. 2000. Human body composition: in vivo methods. *Physiological Reviews* **80**:649–680 DOI 10.1152/physrev.2000.80.2.649.

Fan C, Chen M, Wang X, Wang J, Huang B. 2021. A review on data preprocessing techniques toward efficient and reliable knowledge discovery from building operational data. *Frontiers in Energy Research* **9**:652801 DOI 10.3389/fenrg.2021.652801.

Fan Z, Chiong R, Hu Z, Keivanian F, Chiong F. 2022. Body fat prediction through feature extraction based on anthropometric and laboratory measurements. *PLOS ONE* **17(2)**:e0263333 DOI 10.1371/journal.pone.0263333.

Fan C, Xiao F, Madsen H, Wang D. 2015. Temporal knowledge discovery in big BAS data for building energy management. *Energy and Buildings* **109**:75–89 DOI 10.1016/j.enbuild.2015.09.060.

Figueiredo Filho DB, Júnior JAS, Rocha EC. 2011. What is R2 all about? *Leviathan* **3**:60–68 DOI 10.11606/issn.2237-4485.lev.2011.132282.

Fix E, Hodges JL. 1989. Discriminatory analysis. Nonparametric discrimination: consistency properties. *International Statistical Review/Revue Internationale de Statistique* **57(3)**:238–247 DOI 10.2307/1403797.

Freedman DS, Dietz WH, Srinivasan SR, Berenson GS. 2009. Risk factors and adult body mass index among overweight children: the Bogalusa heart study. *Pediatrics* **123(3)**:750–757 DOI 10.1542/peds.2008-1284.

Friedman JH. 2001. Greedy function approximation: a gradient boosting machine. *Annals of Statistics* **29(5)**:1189–1232 DOI 10.1214/aos/1013203451.

García S, Luengo J, Herrera F. 2015. *Data preprocessing in data mining*. Cham: Springer.

Gheyas IA, Smith LS. 2010. Feature subset selection in large dimensionality domains. *Pattern Recognition* **43(1)**:5–13 DOI 10.1016/j.patcog.2009.06.009.

Gjesdal CG, Halse JI, Eide GE, Brun JG, Tell GS. 2008. Impact of lean mass and fat mass on bone mineral density: the hordaland health study. *Maturitas* **59(2)**:191–200 DOI 10.1016/j.maturitas.2007.11.002.

Glover FW, Kochenberger GA. 2006. *Handbook of metaheuristics*. Cham: Springer Science & Business Media.

Gruzdeva O, Borodkina D, Uchasova E, Dyleva Y, Barbarash O. 2018. Localization of fat depots and cardiovascular risk. *Lipids in Health and Disease* **17(1)**:1–9 DOI 10.1186/s12944-018-0856-8.

Gustavsen A, Grynning S, Arasteh D, Jelle BP, Goudey H. 2011. Key elements of and material performance targets for highly insulating window frames. *Energy and Buildings* **43(10)**:2583–2594 DOI 10.1016/j.enbuild.2011.05.010.

Hastuti J, Kagawa M, Byrne NM, Hills AP. 2018. Anthropometry to assess body fat in Indonesian adults. *Asia Pacific Journal of Clinical Nutrition* **27**:592–598 DOI 10.6133/apjcn.092017.02.

Henry CJ, Shalini D, Ponnalagu O, Bi X, Tan S-Y. 2018. New equations to predict body fat in Asian-Chinese adults using age, height, skinfold thickness, and waist circumference. *Journal of the Academy of Nutrition and Dietetics* **118(7)**:1263–1269 DOI 10.1016/j.jand.2018.02.019.

Hoerl AE, Kennard RW. 1970. Ridge regression: biased estimation for nonorthogonal problems. *Technometrics* **12(1)**:55–67 DOI 10.1080/00401706.1970.10488634.

Holland JH. 1992. Genetic algorithms. *Scientific American* **267(1)**:66–73 DOI 10.1038/scientificamerican0792-66.

Hua J, Tembe WD, Dougherty ER. 2009. Performance of feature-selection methods in the classification of high-dimension data. *Pattern Recognition* **42(3)**:409–424 DOI 10.1016/j.patcog.2008.08.001.

Hussain SA, Cavus N, Sekeroglu B. 2021. Hybrid machine learning model for body fat percentage prediction based on support vector regression and emotional artificial neural networks. *Applied Sciences* **11(21)**:9797 DOI 10.3390/app11219797.

Imai A, Komatsu S, Ohara T, Kamata T, Yoshida J, Miyaji K, Takewa M, Kodama K. 2012. Visceral abdominal fat accumulation predicts the progression of noncalcified coronary plaque. *Atherosclerosis* **222(2)**:524–529 DOI 10.1016/j.atherosclerosis.2012.03.018.

Imran M, Alsuhaibani SA. 2019. A neuro-fuzzy inference model for diabetic retinopathy classification. In: *Intelligent Data Analysis for Biomedical Applications*. Amsterdam: Elsevier, 147–172.

Jain A, Nandakumar K, Ross A. 2005. Score normalization in multimodal biometric systems. *Pattern Recognition* **38(12)**:2270–2285 DOI 10.1016/j.patcog.2005.01.012.

Johnson RW. 1996. Fitting percentage of body fat to simple body measurements. *Journal of Statistics Education* **4(1)**:415 DOI 10.1080/10691898.1996.11910505.

Kalousis A, Prados J, Hilario M. 2007. Stability of feature selection algorithms: a study on high-dimensional spaces. *Knowledge Information Systems* **12(1)**:95–116 DOI 10.1007/s10115-006-0040-8.

Karaman R, Odabas MS, Turkay C. 2024. Estimation of Mung Bean [Vigna radiata (L.) Wilczek] pod shell rate using curve fitting and artificial neural network techniques. *Brazilian Archives of Biology and Technology* **67(36)**:e24230283 DOI 10.1590/1678-4324-2024230283.

Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu T-Y. 2017. LightGBM: a highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems* **30**:3146–3154 DOI 10.5555/3294996.3295074.

Keivanian F, Chiong R, Fan Z. 2023. A fuzzy adaptive evolutionary-based feature selection and machine learning framework for single and multi-objective body fat prediction. ArXiv DOI 10.48550/arXiv.2303.11949.

Kohavi R, John GH. 1997. Wrappers for feature subset selection. *Artificial Intelligence* **97(1–2)**:273–324 DOI 10.1016/s0004-3702(97)00043-x.

Kopelman P. 2007. Health risks associated with overweight and obesity. *Obesity Reviews* **8**:13–17 DOI 10.1111/j.1467-789x.2007.00311.x.

Krishna K, Murty MN. 1999. Genetic K-means algorithm. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* **29(3)**:433–439 DOI 10.1109/3477.764879.

Kumar V, Minz S. 2014. Feature selection: a literature review. *SmartCR* **4(3)**:211–229.

Lai C-M, Chiu C-C, Shih Y-C, Huang H-P. 2022. A hybrid feature selection algorithm using simplified swarm optimization for body fat prediction. *Computer Methods and Programs in Biomedicine* **226(1)**:107183 DOI 10.1016/j.cmpb.2022.107183.

Langley P. 1994. Selection of relevant features in machine learning. In: *Proceedings of the AAAI Fall Symposium on Relevance*, 245–271.

Lichtner-Bajjaoui A. 2021. A mathematical introduction to neural networks. *Available at https://hdl.handle.net/2445/180441*.

Liu H, Motoda H. 1998. *Feature extraction, construction and selection: a data mining perspective*. Berlin, Germany: Springer Science & Business Media.

Liu H, Motoda H. 2007. *Computational methods of feature selection*. Boca Raton, Florida: CRC Press.

Liu H, Motoda H. 2013. *Instance selection and construction for data mining*. Berlin, Germany: Springer Science & Business Media.

Liu H, Yu L. 2005. Toward integrating feature selection algorithms for classification and clustering. *IEEE Transactions on Knowledge Data Engineering* **17(4)**:491–502 DOI 10.1109/tkde.2005.66.

Luengo J, García S, Herrera F. 2012. On the choice of the best imputation methods for missing values considering three groups of classification methods. *Knowledge and Information Systems* **32(1)**:77–108 DOI 10.1007/s10115-011-0424-2.

Maingi MN. 2015. Survey on data preprocessing concept application data mining. *International Journal of Mining Science and Technology* **4**:1901–1902 DOI 10.21275/SUB151633.

McCulloch WS, Pitts W. 1943. A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics* **5(4)**:115–133 DOI 10.1007/bf02478259.

Mlambo N, Cheruiyot WK, Kimwele MW. 2016. A survey and comparative study of filter and wrapper feature selection techniques. *The International Journal of Engineering and Science (IJES)* **5(8)**:57–67.

Muni DP, Pal NR, Das J. 2006. Genetic programming for simultaneous feature selection and classifier design. *IEEE Transactions on Systems, Man, Cybernetics* **36(1)**:106–117 DOI 10.1109/tsmcb.2005.854499.

Ólafsson S. 2006. Chapter 21 metaheuristics. In: Henderson SG, Nelson BL, eds. *Handbooks in Operations Research and Management Science*. United States: Elsevier, 633–654.

Opitz D, Maclin R. 1999. Popular ensemble methods: an empirical study. *Journal of Artificial Intelligence Research* **11**:169–198 DOI 10.1613/jair.614.

Pagan AR, Hall AD. 1983. Diagnostic tests as residual analysis. *Econometric Reviews* **2(2)**:159–218 DOI 10.1080/07311768308800039.

Parvandeh S, Yeh H-W, Paulus MP, McKinney BA. 2020. Consensus features nested cross-validation. *Bioinformatics* **36**:3093–3098 DOI 10.1101/2019.12.31.891895.

Patro S, Sahu KK. 2015. Normalization: a preprocessing stage. ArXiv DOI 10.48550/arXiv.1503.06462.

Peng H, Fan Y. 2015. Direct l_ (2, p)-norm learning for feature selection. ArXiv DOI 10.48550/arXiv.1504.00430.

Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. 2018. CatBoost: unbiased boosting with categorical features. In: *Advances in Neural Information Processing Systems.* Vol. 31.

Ramchandran A, Sangaiah AK. 2018. Unsupervised anomaly detection for high dimensional data—an exploratory analysis. In: *Computational Intelligence for Multimedia Big Data on the Cloud with Engineering Applications*. Amsterdam, Netherlands: Elsevier, 233–251.

Rostami M, Moradi P. 2014. A clustering based genetic algorithm for feature selection. In: *2014 6th Conference on Information and Knowledge Technology (IKT)*. Piscataway: IEEE, 112–116.

Ruiz R, Riquelme JC, Aguilar-Ruiz JS. 2005. Heuristic search over a ranking for feature selection. In: *International Work-Conference on Artificial Neural Networks*. Cham: Springer, 742–749.

Salamunes ACC, Stadnik AMW, Neves EB. 2018. Estimation of female body fat percentage based on body circumferences. *Revista Brasileira de Medicina do Esporte* **24(2)**:97–101 DOI 10.1590/1517-869220182402181175.

Santos D. 2023. Predicting body fat percentage: a machine learning approach. DOI 10.20944/preprints202310.0929.v1.

Sengupta RN, Gupta A, Dutta J. 2016. *Decision sciences: theory and practice*. Boca Raton: Crc Press.

Shao YE. 2014. Body fat percentage prediction using intelligent hybrid approaches. *The Scientific World Journal* **2014(1)**:1–8 DOI 10.1155/2014/383910.

Sinaga KP, Yang M-S. 2020. Unsupervised K-means clustering algorithm. *IEEE Access* **8**:80716–80727 DOI 10.1109/access.2020.2988796.

Svendsen OL, Haarbo J, Heitmann BL, Gotfredsen A, Christiansen C. 1991. Measurement of body fat in elderly subjects by dual-energy x-ray absorptiometry, bioelectrical impedance, and anthropometry. *The American Journal of Clinical Nutrition* **53(5)**:1117–1123 DOI 10.1093/ajcn/53.5.1117.

Thorpe M. 1988. *Handbook of education technology*. Ellington, Percival Race.

Tibshirani R. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* **58(1)**:267–288 DOI 10.1111/j.2517-6161.1996.tb02080.x.

Uçar MK, Ucar Z, Köksal F, Daldal N. 2021. Estimation of body fat percentage using hybrid machine learning algorithms. *Measurement* **167(2)**:108173 DOI 10.1016/j.measurement.2020.108173.

Wang YC, McPherson K, Marsh T, Gortmaker SL, Brown M. 2011. Health and economic burden of the projected obesity trends in the USA and the UK. *The Lancet* **378(9793)**:815–825 DOI 10.1016/s0140-6736(11)60814-3.

Wang L, Ni H, Yang R, Pappu V, Fenn MB, Pardalos PM. 2014. Feature selection based on meta-heuristics for biomedicine. *Optimization Methods and Software* **29(4)**:703–719 DOI 10.1080/10556788.2013.834900.

Wang J, Su X. 2011. An improved K-Means clustering algorithm. In: *2011 IEEE 3rd International Conference on Communication Software and Networks*. Piscataway: IEEE, 44–46.

Willmott CJ, Matsuura K. 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research* **30**:79–82 DOI 10.3354/cr030079.

Wold S, Sjöström M, Eriksson L. 2001. PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems* **58(2)**:109–130 DOI 10.1016/s0169-7439(01)00155-1.

World Health Organization. 2018. Noncommunicable diseases country profiles 2018. *Available at https://www.who.int/publications/i/item/9789241514620*.

Xu S, Nianogo RA, Jaga S, Arah OA. 2023. Development and validation of a prediction equation for body fat percentage from measured BMI: a supervised machine learning approach. *Scientific Reports* **13(1)**:8010 DOI 10.1038/s41598-023-33914-5.

Yang X-S. 2010. *Engineering optimization: an introduction with metaheuristic applications*. Hoboken, New Jersey: John Wiley & Sons.

Yusta SC. 2009. Different metaheuristic strategies to solve the feature selection problem. *Pattern Recognition Letters* **30(5)**:525–534 DOI 10.1016/j.patrec.2008.11.012.

Zeng X, Chen Y-W, Tao C, van Alphen D. 2009. Feature selection using recursive feature elimination for handwritten digit recognition. In: *2009 Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*. Piscataway: IEEE, 1205–1208.

Zhang H, Ho T-B, Lin M-S. 2004. An evolutionary K-means algorithm for clustering time series data. In: *Proceedings of 2004 International Conference on Machine Learning and Cybernetics (IEEE Cat No 04EX826)*. Piscataway: IEEE, 1282–1287.

Zhang S, Zhang C, Yang Q. 2003. Data preparation for data mining. *Applied Artificial Intelligence* **17(5–6)**:375–381 DOI 10.1080/713827180.

Zheng A, Casari A. 2018. *Feature engineering for machine learning: principles and techniques for data scientists*. Sebastopol, California: O'Reilly Media, Inc.

Zou H, Hastie T. 2005. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67(2)**:301–320 DOI 10.1111/j.1467-9868.2005.00503.x.