An Framework for Knee Osteoporosis Classification Using Hybrid Multi-Modal Feature Fusion and Fuzzy Harmony Search Optimization

Saud M. Abdul Razzaq and Baydaa I. Khaleel
Department of Computer Science, College of Computer Science and Mathematics, University of Mosul, 41002 Mosul, Iraq
saud.23csp48@student.uomosul.edu.iq, baydaaibraheem@uomosul.edu.iq

Keywords: X-Ray Images, Osteoporosis, Machine Learning, Harmony Search Algorithm, Gath-Geva Algorithm.

Abstract: Knee osteoporosis is a structural disorder characterized by decreased bone mineral density, which increases the risk of fractures and significant morbidity. Early and accurate diagnosis is crucial for timely intervention and treatment, as traditional diagnostic methods rely on manual interpretation of X-ray images, which is time-consuming and subject to inter-examiner variability. This research presents an automated system for classifying and detecting knee osteoporosis that leverages a novel methodology for extracting, selecting, and classifying features via machine learning. We proposed a new hybrid multi-modal feature fusion technique, Hybrid Multi-Modal Feature Fusion (HMMFF), that integrates traditional feature extraction methods such as GLCM, HOG, and ORB with deep learning features from models including YOLOv11, VGG-16, and ResNet50. To improve the classification efficiency of the proposed system, feature selection was performed using the harmony search optimization algorithm. We also proposed a new hybrid feature selection method, the Hybrid Fuzzy Harmony Search Algorithm (HFHSA), that combines harmony search with Gath-Geva fuzzy clustering. The selected features were then fed into the proposed machine learning classifiers and evaluated. The results indicated that the XGBoost model achieved the highest accuracy rate, reaching 93.58% when using the Harmony Search algorithm alone for feature selection, While the Gradient Boosting model achieved the highest accuracy compared to the other models rate rose to 94.87% when using the HFHSA technique for feature selection. These results confirm the effectiveness of applying the HMMFF technique for feature extraction, along with feature selection using the HFHSA technique, in identifying important features, thereby providing an accurate diagnostic tool that enhances early detection of osteoporosis.

1 INTRODUCTION

Osteoporosis is a major public health problem affecting millions of people worldwide, especially the elderly. Osteoporosis is a systemic skeletal disease characterized by decreased bone mass and deterioration of bone microstructure, leading to increased bone fragility and a higher risk of fractures. Osteoporosis is diagnosed based on bone mineral density (BMD) measurements; a T-score of less than -2.5 standard deviations indicates the condition [1], [2].

Although osteoporosis in the knee does not occur as frequently as in other skeletal organs, including the hip or spine, it is a major clinical issue because of its effect on mobility. It may result in chronic pain and mobility issues, as well as low quality of life. Knee is a weight bearing component of the body and plays a critical role in movement and day to day activities. Osteoporosis may reduce the capacity of an individual to engage in any physical activity which results in sedentary lives and other health related issues. As such, proper clinical management and prevention of complications related to fractures require early identification and proper classification [3], [4].

Conventional options of osteoporosis diagnosis are subject to various difficulties as the changes in the bone density are subtle and require special knowledge, which may result in different views and human error. These problems have led to more use of artificial intelligence. Artificial intelligence and machine learning have transformed the medical image analysis process and provided reproducible, automatic, and objective solutions to the diagnostic task. The recent progress in computer vision and deep learning has shown significant success in several medical imaging tasks such as identifying and diagnosing osteoporosis using X-ray images. Such automated systems have been shown to break down image patterns which are complicated, finding pertinent features and classifying the level of disease with accuracy in as high as human experts [5], [6].

In this paper, we proposed a methodology of the automated detection of knee osteoporosis and the efficacy of several machine learning methods to detect the disease correctly. It also provides a number of significant contributions:

- An innovative method of HMMFF feature extraction is proposed, that is, combining the traditional methods (GLCM, HOG, ORB) with deep learning-based features of several models (YOLOv11, VGG-16, ResNet50), which allows obtaining a variety of comprehensive features.
- A novel feature selection method is proposed using a hybrid HFHSA technique that integrates harmony search optimization with Gath-Geva fuzzy clustering. This method leverages the comprehensive search capabilities of harmony search and the uncertainty management of fuzzy logic to identify optimal feature sets. This can improve the identification of important data related to bone density, thereby increasing the performance efficiency of classifiers.

2 RELATED WORK

Applications and research for detecting and classifying osteoporosis have seen significant development in recent years through the application of artificial intelligence techniques. Early methods relied mainly on machine learning and deep learning techniques, and these works are listed below.

In 2023, Kumar et al. proposed a model for classifying osteoporosis knee x-ray datasets, an ensemble-based classifier model was built using three convolutional neural network (CNN) architectures: Inception v3, Xception, and ResNet 18, taking into account two different factors in the decision scores generated by the previously described base classifiers, the proposed ensemble technique uses a rank-based fuzzy consolidation of the classifiers, unlike simple consolidation procedures documented in the literature, the proposed ensemble approach considers the confidence level of the base classifiers' recommendations when making final predictions on test samples, a publicly available benchmark dataset was used to evaluate the constructed framework using a five-fold cross-validation technique, with a loss of 0.082, the proposed model achieved an accuracy of 93.5% [7].

In 2023, Wani and Arora proposed a method for diagnosing osteoporosis from knee X-ray images using transfer learning in convolutional neural networks, such as ResNet-18, VggNet-16, AlexNet, and VggNet-19, the results show that AlexNet achieved the best performance with an accuracy of 91%, while VggNet-19 achieved the lowest accuracy with an accuracy of 84.2%, The study suggests that a deep learning system with transfer learning can aid doctors in detecting osteoporosis early, thereby reducing the risk of fractures[4].

In 2023, Ponni et al. presented a study analyzing medical images and clinical data to evaluate the effectiveness of various deep learning models, including convolutional neural networks (CNNs), artificial neural networks (ANNs), multilayer perceptrons (MLPs), and long-term short-term memory networks (LSTMs). Critical performance metrics, including accuracy, loss, sensitivity, receiver operating characteristics (ROC), area under the curve (AUC), and scalability, were used to assess the effectiveness of each method. The results indicated that CNNs demonstrated remarkable accuracy in both diagnoses. In diagnosing osteoporosis, CNNs achieved an impressive accuracy of 94.3%, significantly outperforming artificial neural networks (ANNs) (87%), artificial neural networks (MLPs) (80%), and long-term short-term neural networks (LSTMs) (74%). In diagnosing osteoporosis, CNN technology demonstrated superior accuracy at 99%, followed by ANN at 88.7%, MLP at 87%, and LSTM at 80.3%. This comprehensive research clearly indicates that deep learning has significant potential to transform the diagnosis of osteoporosis [8].

As part of 2023, Dodamani and Ajit provided a comparative study to assess the accuracy and efficiency of the modern deep learning CNN models, VGG16, VGG19, DenseNet121, ResNet50, and InceptionV3, to detect and classify the normal and osteoporotic conditions, the study used 830 X-ray images of the spine, hand, leg, knee, and hip, with the necessary results showing that DenseNet121 was an outstanding model with an accuracy rate of 93.4[9].

In 2024, Shen studied the diagnosis of osteoporosis without requiring any additional test besides knee x-ray analysis based on deep learning algorithms, the strategy relied on the artificial intelligence technology to enhance the diagnostic accuracy and efficiency as it offers a convenient and non-invasive method of diagnosing osteoporosis features in knee x-rays, the VGG 19 model, a model with the ability to extract complex features of 2D images, was utilized, the model with an 89% accuracy using both Kaggle and Mendeley datasets, was used [10].

Elsewhere in 2024, Naguib et al. carried out a study according to which the images of knee X-rays were classified in three groups, such as normal, osteoporosis, and osteopenia. The project experienced the use of deep learning and overflow mechanism which put forward two different kinds of blocks. This proposed model was trained, tested and validated on two knee datasets and two pre-trained models ResNet50 and AlexNet were subjected to transfer learning. The findings were compared to the proposed model that had better performance as compared to the pre-trained models. The Superfluity DL was the most accurate of all pre-trained models with 85.42% and 79.39% on dataset 1 and 2, respectively [11].

In 2025, OM, A., & Gunasundari, R. created an innovative hybrid deep learning algorithm that combines a ResNet50 network in space feature extraction with Gateway Repetition Units (GRUs) in learning time dependency, which allows analyzing X-ray images on the knee better. The model had a classification accuracy of 95.65 and was able to differentiate between osteoporosis and normal conditions and made itself as a good non-invasive screening method of the early detection of osteoporosis [12].

Muhammad et al. presented a deep learning hybrid model to detect osteoporosis in 2025 based on the KOP dataset. This method is a combination of two pre-trained models, DenseNet169 and ViT, using a custom attention model. The classification performance was also greatly enhanced since it effectively obtained spatial and channel specific information with high accuracy (0.8611), high specificity (0.9474) and high precision (0.9286). The findings indicate that the model can assist in improving the accuracy and efficiency of osteoporosis diagnosis through studying the X-ray images of the knee, which contributes to the possibility of clinical use to improve the early detection and specific interventions. This study emphasizes that the adoption of advanced models is important to enhance bone characterization and creates the prospects of a further study into more various bone-related diseases [13].

Qureshi et al. (2023) introduced a deep learning based on the ResNet-50 architecture to detect osteoporosis in knee X-ray images as a computational method. A transfer learning approach leads to better performance of the model by using ResNet-50 pre-trained weights. The approach will be to generate an image dataset that is further subdivided into training and test sets to determine the accuracy of the classifier. The model is very effective in terms of accuracy, sensitivity, and specificity with a maximum of 90. The importance of early detection techniques in preventing fractures and possible cost reduction of osteoporosis treatment form a vital part of this study [14].

3 METHODOLOGY

The goal of this methodology is to make the diagnosis and classification of osteoporosis more accurate and effective with the use of the suggested model. It does this through the use of smart methods that are able to detect the constraints of the conventional means of diagnostic methods. The objective of the process is to categorise typical cases of osteoporosis cases such divided into the following categories: normal, osteopenia, osteoporosis. It is achieved through the application of the different machine learning classification models and their comparison.

Data was prepared by collecting and enriching it to enhance model performance. The dataset, collected from the Kaggle platform, consisted of 1947 knee X-ray images divided into training and test sets. The data were then balanced using the RandomOverSampler to verify the distribution of categories within the dataset and correct any imbalances, achieving a largely balanced distribution.

Figure 1 illustrates the classification of osteoporosis in the knee. Table 1 shows the number of samples in the classification.

Figure 1: Classification of the of osteoporosis of the knee.

Table 1: Number of samples in classification.

Class	Total
Normal	780
Osteopenia	374
Osteoporosis	793

3.1 Image Preprocessing

Image processing is crucial for improving the quality of images input into the system. It prepares the data effectively and efficiently for subsequent stages, thereby contributing to image recognition and differentiation. The image processing was carried out through a series of steps applied to X-ray images to simplify the computational complexity of handling them, as follows:

- Applying Contrast Limited Adaptive Histogram Equalization (CLAHE) to improve contrast, address uneven lighting in the image, and enhance the quality of images with blurred details.
- Applying a Gaussian Blur to reduce noise in the image.
- Improving image clarity and highlighting important details through Sharpening.
- Applying the Morphological Gradient to define and highlight edges in the images to reveal the external boundaries of bones.
- Resizing all images to 128 x 128 pixels to ensure uniformity and consistent input dimensions during the feature extraction stage.

Figure 2 shows the series of preprocessing steps that were applied to the X-ray images of the inserted knee.

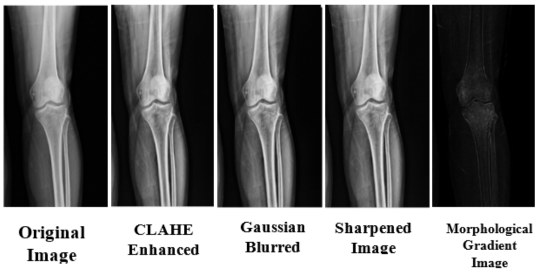

Figure 2: Steps for pre-processing images.

3.2 Feature Extraction

Feature extraction is a crucial process for data analysis and classification, contributing significantly to model efficiency and error reduction. It involves analyzing and understanding data to identify the most important features that contribute to achieving the desired outcome. The algorithms used to accomplish their tasks whether object recognition, classification, or diagnosis often rely on extracted features. The presence of irrelevant or redundant features can lead to problems with algorithm functionality and execution efficiency, as it expands the search space for the problem being solved, thereby increasing the data volume and the time required to complete the task [15], [16].

3.2.1 The HMMFF Feature Extraction Method:

Our research proposed a novel feature-extraction method, Hybrid Multi-Modal Feature Fusion (HMMFF). This method relies on a comprehensive, multi-level approach to feature extraction and fusion. The aim is to leverage the complementary capabilities of each feature set. It incorporates techniques such as HOG, GLCM, and ORB to obtain local and geometric pattern-based features, respectively, while VGG16, YOLOv11, and ResNet50 provide more abstract, hierarchical representations learned from large datasets. The features extracted from each method are then combined and standardized to obtain an effective image representation. These features are then selected and fed into machine learning models for classification, as follows:

1) A-Feature Extraction Stage Using Traditional Techniques:
   - Gray Level Concurrency Matrix (GLCM). GLCM is used to analyze images based on texture analysis. It reveals how grayscale is distributed in an image by counting how often a pair of grayscale values occurs at a given location [17]. GLCM features are extracted at a distance of 1 and at angles of 0, 45, 90, and 135 degrees, using 256 grayscale levels. Five texture attributes are then calculated, including contrast, variation, homogeneity, energy, and correlation, to characterize the spatial relationships between pixel densities.
   - Oriented Gradient Histogram (HOG). This thesis utilizes HOG, a computer vision algorithm used to extract features from images by detecting objects within them [18]. It supports the extraction of local information from within the image, employing cells with dimensions of 16 × 16 pixels and a block size of 1 × 1.
   - Oriented FAST and Rotated Brief (ORB). ORB is used to identify, recognize, and describe key points in the image for feature extraction. These features include rotation and stability under varying lighting conditions [19]. ORB descriptions are generated for up to 500 key points to ensure consistent dimensions and a flat, localized representation of stable features.

2) B- The feature extraction phase using deep learning models:

Pre-trained deep learning models were used to extract highly accurate features by disabling the Fully Connected layer, which is responsible for classification. Deep convolutional features were extracted from the VGG-16 model, pre-trained on ImageNet. This model is a 16-layer deep neural network known for its simplicity and efficiency in representing high-dimensional images with a recurrent structure of convolutional and clustered layers [20]. It was used to extract deep representations from radiographic images, yielding a 512-dimensional vector of high-level features. Feature extraction was also performed using the YOLOv11 model, known for its fast, accurate object detection, yielding a 128-dimensional feature vector. Additionally, the pre-trained ResNet50 model was used. This is a 50-layer deep network that uses residual connections, allowing it to be trained without vanishing gradients. This enables access to more complex representations and is more effective in recognizing subtle patterns in medical images, yielding a feature vector containing 2,048 features [20].

3) Feature Combination Phase. The features extracted using the previous methods were grouped and combined to obtain a comprehensive vector containing 4,385 features. These features represent low-level texture information, geometric patterns, local key point descriptors, and high-level semantic attributes. This created a robust and discriminatory feature space suitable for subsequent classification tasks. This combined vector was then applied, after feature selection, as input to subsequent classification models. Figure 3 illustrates the HMMFF feature extraction process.

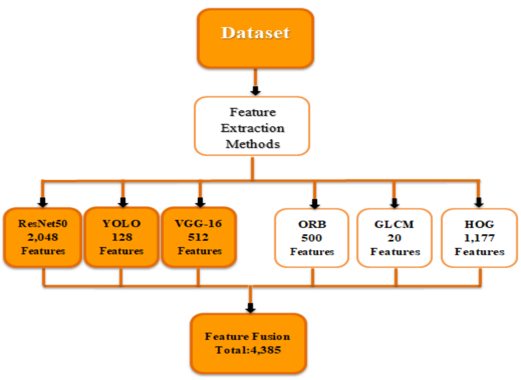

Figure 3: Extracting features using the HMMFF method.

3.3 Feature Selection

Choosing features is also a major challenge to the machine learning methods. It tries to find a few very relevant and powerful features in the original data, which results in better performance and accuracy of machine learning models at lower computational costs. The methods of feature extraction and selection are employed to improve the effectiveness of the model [21]. The feature extraction process converts the original data into new variables that capture the necessary information, and the feature selection finds and depicts simpler and easier to understand models of the relationship between variables and outcomes, thus simplifying the effect of computation and speeding up the training process [22].

3.3.1 Hybrid Fuzzy Harmony Search Algorithm (HFHSA)

The new hybrid HFHSA algorithm applied in this work, an integration of the Gath-Geva algorithm and harmony search algorithm, is a new method of optimization of the performance of algorithms in feature selection tasks, as this approach can explore the solution space more thoroughly, as well as avoid dependence on predefined parameters and overcome the shortcomings of used evaluation functions in feature search algorithms, most of the functions pay attention only to a single metric, which is accuracy, and as a result, may be subjected to the risk of falling into a local solution, which can be improved through a more extensive exploration of the solution space,
Harmony search algorithm, first proposed in 2001 by Korean scientist Zong Woo Geem, is a metaheuristic algorithm inspired by musical improvisation, it relies on a foundation consisting of the harmony memory update equation, the harmony memory acceptance rate update equation, and the use of the pitch modulation rate to determine the optimal solution, by using these mechanisms, the algorithm seeks to navigate the search space efficiently, adapt to potential changes in the optimization problem, and converge to the global optimum while avoiding falling into it, in local optimization [23], [24]. It is used to address complex problems through intelligent exploration of the solution space, the stochastic nature of the algorithm facilitates in-depth exploration of the search space and reduces the likelihood of falling into the trap of local minima, this property enables the harmony search method to surpass traditional methods, this method is also characterized by its insensitivity to initial settings [25].

The Harmony search algorithm is used to explore the space of possible solutions, while fuzzy logic is applied to evaluate the quality of these solutions, this is achieved by enhancing the fitness function by leveraging the strengths of each algorithm to enhance the flexibility and accuracy of the evaluation process, this is achieved by replacing traditional fitness functions with a multi-criteria fuzzy inference system, the quality of the solution (fitness) is evaluated based on three metrics: classification accuracy, feature stability, and redundancy between features, these values are converted into fuzzy metrics using the Gath-Geva model, which is based on mixed Gaussian distributions.

The evaluation process begins by calculating and finding the numerical sharp values for the specified metrics and for each solution (a set of features) in the harmony memory, membership functions derived from these are then used, the Gath-Geva algorithm converts these numeric values into fuzzy values, this is done according to the following steps:

3.3.1.1 Calculating Basic Metrics

First, a clear numerical value is calculated for each of the three metrics for the proposed feature set, which represents a proposed solution in the Harmony algorithm, as follows:

1) Calculating accuracy. This metric measures the model's ability to correctly classify using the selected set of features.

   The standard mathematical equation for calculating accuracy is:

$$\text{Accuracy} = \frac{\text{Number of correct Prediction}}{\text{Total Number of samples}} \quad (1)$$

2) Stability Calculation. Measures the stability of an individual feature in terms of its predictive ability and consistency, as it does not contain much noise or random fluctuations, stability is calculated by calculating two sub-measures:
   - Variance-dependent stability: Measures the ratio of the variance of the feature to the mean, a feature is more stable when its variance is low compared to its mean.
   - Entropy-dependent stability: Measures the degree of randomness in the distribution of values for a feature, a feature is more stable when its entropy is low. The following mathematical equation is used to calculate stability [26]:

$$\text{stability} = 0.6 \times \frac{1}{1+\frac{\text{Var}(X)}{|\text{Mean}(X)|+\epsilon}} + 0.4 \times \left(1 - \frac{\text{H}(X)}{\log(N)}\right). \quad (2)$$

Where stability represents an indicator of the stability of the feature, Var(X) is the variance in the values of the feature, ϵ is a small value to avoid dividing by zero, H(X) is the entropy representing a measure of the randomness of the values in the feature, and N is the number of samples.

3) Redundancy. It measures the degree of correlation between selected features, The goal is to select a set of features that provides diverse, non-redundant information.

The standard mathematical equation for average correlation coefficients is [26]:

$$\text{Redundancy} = \frac{2}{k(k-1)} \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} |\rho_{ij}|. \quad (3)$$

Where:
- Redundancy represents an index of redundancy in the selected features;
- k represents the selected features;
- ρ_{ij} is the correlation coefficient between feature i and feature j.

3.3.1.2 Converting Metrics to Fuzzy Values

To convert numerical values for solution evaluations into fuzzy values, the Gath-Geva algorithm is first trained on the training data, The model divides the feature space into a set of clusters, each with its own center, covariance matrix, and Gaussian distribution, The Gath-Geva algorithm is then used to generate a fuzzy affiliation for each metric, representing each metric as a multi-level fuzzy variable, and calculating the non-fuzzy value affiliation score for each level.

Calculating the Membership Score. Each metric value is treated as a new data point, The values are normalized, and then their membership scores are calculated for each cluster.

The basic equation for calculating the membership score of point xi in cluster Cj in the Gath-Geva algorithm is [27], [28]:

$$p_i(x_k) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_i|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x_k - \mu_i)^T \Sigma_i^{-1}(x_k - \mu_i)\right). \quad (4)$$

Where:
- $p_i(x_k)$ is the degree of belonging of point x_k to cluster i;
- d is the number of features;
- $|\Sigma_i|$ is the determinant of the covariance matrix of cluster i;
- μ_i represents the center of cluster i;
- Σ_i is the covariance matrix of cluster i;
- x_k represents the value vector of point k.

Then, the appropriate cluster is chosen for each of the three measures, where the highest values of belonging scores are selected for the accuracy and stability measures, while the lowest values of belonging scores are chosen for the repetition measure, given that the lower the repetition, the better the results, The measures are then combined after converting them to fuzzy values, using the following fuzzy inference system to obtain the final fitness value that guides the search process in the Harmony Search algorithm:

$$\text{Fitness} = (\text{Accuracy}_f)^{w_1} \times (\text{Stability}_f)^{w_2} \times (1 - \text{Redundancy}_f)^{w_3} \quad (5)$$

Where w1, w2, w3 are weights used to determine the importance of each metric in the final evaluation.

Figure 4 shows the scheme of the new hybrid HFHSA method, The proposed hybrid method for feature selection demonstrates that incorporating fuzzy logic into the fitness calculation stages enhances the algorithm's ability to handle non-linear complexities in the data, reduces reliance on sharp values, and improves the efficiency of feature selection by increasing stability and reducing redundancy while maintaining predictive performance, after applying the HFHSA method to the features extracted using the HMMFF method, 420 features were selected out of a total of 4,385 features, Thus, the approach followed in the study helps in discovering an ideal subset of features by reducing the complexity of the resulting set of features after merging, which will be used in machine learning models for classification.

3.4 Proposed Algorithm

Classification in medical imaging is a difficult and complex challenge, but it is crucial for the early diagnosis and detection of diseases, however, the continuous development of machine learning algorithms has contributed to improving the classification process, in the proposed model for diagnosing knee osteoporosis, several machine learning algorithms were employed. The classification process of the input images is carried out after basic operational procedures, including data balancing to correct imbalances in datasets, thereby mitigating bias in algorithmic predictions, as well as preprocessing the input images to facilitate feature extraction. In addition, the feature extraction process is implemented using the (HMMFF) feature extraction technique, which is implemented using traditional techniques such as GLCM, HOG, and ORB, along with deep learning models such as YOLO11, VGG-16, and ResNet50 to extract high-resolution features, with the fully connected layer responsible for classification within the model turned off, after that, the extracted features are combined into two sets to produce a comprehensive and effective representation of the image, resulting in 4,385 derived features for each image, totaling for all 1,947 images, The combined feature selection is then performed in two ways using an algorithmic method, harmony search where the total number of features that were selected was (2154) features from the total number of features and then entered to conduct classification and then the hybrid (HFHSA) technique was also used to select the features by identifying the most influential features in the classification process where the features selected in this method were (420) which will also be entered into the different machine learning models to conduct a multi-classification of knee bone cases and then compare the results of the different machine learning models applied, as shown in Figure 5, which illustrates the proposed methodology.

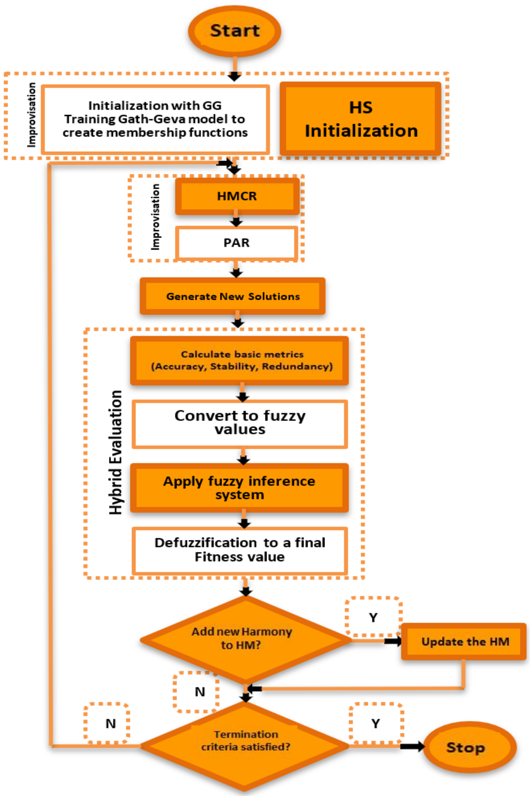

Figure 4: Schematic diagram of the new hybrid HFHSA method.

3.5 Results and Discussion

This research proposes the development of an intelligent system for detecting and classifying knee osteoporosis into three categories: normal, osteopenia, and osteoporosis. The proposed model was implemented using Python to perform detection and classification based on knee X-ray images. All proposed models were trained and evaluated using several established benchmarks, and their performance was then compared to determine the most accurate and efficient model for detecting osteoporosis.

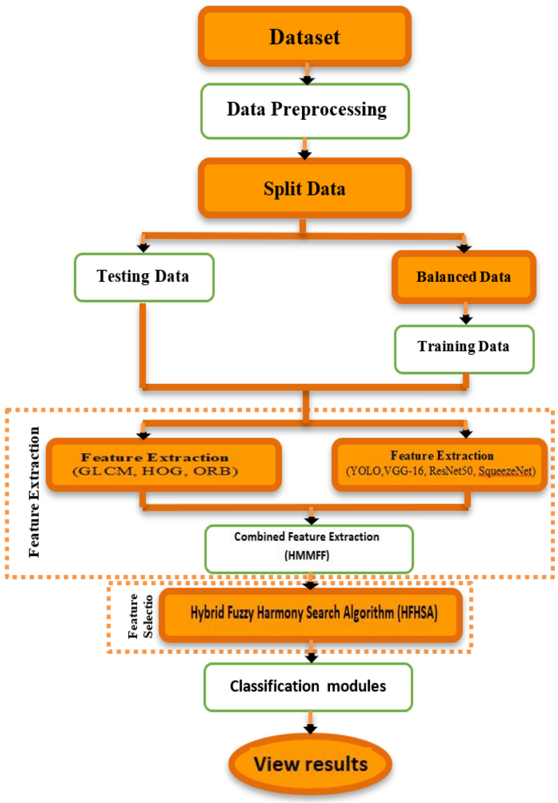

Figure 5: Flowchart of the proposed method of work.

The proposed classification system begins with a comparative methodology aimed at evaluating the performance of two feature selection techniques and their impact on the classification accuracy of fragility images. The results in Tables 2 and 3 show an improvement in the performance of most classifiers when using the feature selection set via HFHSA. This experimentally demonstrates the effectiveness of the hybridization strategy in enhancing the quality of the selected features. The XGBoost model achieved the best performance the two techniques compared to the other classifiers, with an accuracy rate of 93.59% and an F1 coefficient of 92.91% when using HAS. This was followed by ExtraTrees at 93.33% and GradientBoosting at 93.08%. DecisionTree and Bagging showed the lowest performance, with accuracy rates of 91.8%, 91.3%. This demonstrates the superiority of ensemble models based on progressive boosting in medical classification tasks that rely on multi-source features.

When switching to the HFHSA application, there was a clear improvement in the performance of all models. GradientBoosting's accuracy increased from 93.08% to 94.87%, topping the results with an improvement of 1.79%. XGBoost's accuracy also increased from 93.59% to 94.1% with an improvement of 0.51%. ExtraTrees improved from 93.33% to 93.85%, and RandomForest saw an increase, achieving 93.59%. Bagging achieved the greatest performance improvement with an improvement of 2.05%, as its accuracy increased from 91.28% to 93.33%, which is the largest relative improvement among all models. In contrast, DecisionTree achieved a significant improvement of 93.08%. The overall pattern of these improvements is attributed to HFHSA's ability to expand the search space by combining HSA's random exploration with the flexibility of fuzzy logic in handling uncertainty, which reduces redundancy in selected features and enhances the models' ability to generalize. Table 2 shows the performance results of machine learning models with the set of features selected using the harmony search optimization algorithm. Table 3 also shows the performance results of machine learning models with the set of features selected using the new hybrid (HFHSA) technique. Table 4 shows the detailed performance results of machine learning models according to each disease category, with a set of features selected using the HFHSA technique, while.

Table 2: Performance results of the proposed machine learning models with feature selection using harmony search algorithm method.

Classifier	Accuracy	Precision	Recall	F1
RandomForest	0.91794	0.90844	0.906884	0.907523
ExtraTree	0.93333	0.92608	0.924240	0.924880
GradientBoosting	0.93076	0.91823	0.929026	0.923077
Bagging	0.91282	0.90030	0.914311	0.906230
DecisionTree	0.91794	0.90806	0.911580	0.909252
XGBoost	0.93589	0.92568	0.933340	0.929054

Table 3: Performance results of the proposed machine learning models with feature selection using the HFHSA method.

Classifier	Accuracy	Precision	Recall	F1
RandomForest	0.935897	0.932057	0.930750	0.931091
ExtraTree	0.938462	0.933532	0.932927	0.933128
GradientBoosting	0.948718	0.939950	0.941393	0.940576
Bagging	0.933333	0.923077	0.928774	0.925783
DecisionTree	0.930769	0.919487	0.926638	0.922830
XGBoost	0.941026	0.933750	0.935023	0.934240

Table 4: Detailed performance results of machine learning models according to each disease category.

Proposed model	Disease Class	Precision	Recall	F1-score
Random Forest	Normal	0.98	0.94	0.96
	Osteopenia	0.91	0.91	0.91
	Osteoporosis	0.91	0.95	0.93
ExtraTrees	Normal	0.97	0.95	0.96
	Osteopenia	0.91	0.91	0.91
	Osteoporosis	0.92	0.94	0.93
Gradient Boosting	Normal	0.99	0.96	0.97
	Osteopenia	0.89	0.91	0.90
	Osteoporosis	0.94	0.96	0.95
Bagging	Normal	0.96	0.96	0.96
	Osteopenia	0.87	0.91	0.89
	Osteoporosis	0.94	0.92	0.93
Decision Tree	Normal	0.97	0.95	0.96
	Osteopenia	0.86	0.91	0.88
	Osteoporosis	0.93	0.92	0.93
XGBoost	Normal	0.98	0.95	0.96
	Osteopenia	0.89	0.91	0.90
	Osteoporosis	0.93	0.95	0.94

The results analysis in Table 4 provides a detailed picture of performance and metrics such as accuracy, recall, and F1 score for each disease category. A deeper analysis of the results reveals an overall improvement in the performance of the machine learning models used in the study when employing HFHSA technology. However, there is variation in model performance across different disease categories, highlighting the strengths and weaknesses of each model in handling the complexities of classification within these categories.

The GradientBoosting model achieved the highest F1 score for the normal category (0.97) with an exceptional precision score of 0.99, indicating near-perfect predictive accuracy in classifying normal cases. It also achieved F1 scores of 0.90 for osteopenia and 0.95 for osteoporosis, with recall rates of 0.96 for the latter two categories, demonstrating its ability to detect disease cases with a high degree of accuracy. This excellence is attributed to the gradient boosting algorithm's incremental model-building process, minimizing accumulated errors and enabling it to capture subtle boundary lines between categories. In contrast, the XGBoost model demonstrated strong and balanced performance across all three categories, with an F1 coefficient of 0.96 for the normal category, 0.90 for the Osteopenia category, and 0.94 for the Osteoporosis category, along with a high Precision of 0.98 for the normal category. This balance is reflected in XGBoost's ranking as the top-performing model overall within the HSA framework and one of the best models within the HFHSA. RandomForest and ExtraTrees showed remarkably similar performance, both achieving an F1 coefficient of 0.96 for the normal category, while their results were comparable for the Osteopenia (0.91) and Osteoporosis (0.93 for both RandomForest and ExtraTrees). This similarity is explained by the shared basis of the random tree structure between the two models, with a slight difference stemming from the ExtraTrees' split-point selection mechanism, which introduces greater randomness, thus slightly improving generalization. The DecisionTree model achieved an acceptable F1 coefficient of 0.96 for the normal category but recorded the lowest performance among the models in the osteopenia category, with an F1 coefficient of only 0.88. This reveals the limitations of a single tree in handling characteristic overlap between adjacent categories.

The analysis results clearly show that the main challenge is evident in classifying the osteopenia category, indicating the difficulty of differentiating it compared to the other two. The low recall suggests several factors, such as the overlap in imaging characteristics with normal cases and the significant variation in the degree of internal bone deterioration within osteopenia. In normal disease cases, the models exhibited mixed performance, indicating their ability to recognize normal bone characteristics, although some challenges exist in distinguishing between normal and osteopenia categorical boundaries and the overlap between the two. From a clinical perspective, this finding is highly significant, as the low recall rate for the osteopenia category suggests that some transitional cases may be misclassified as normal. Although the medical cost of a false negative diagnosis typically outweighs that of a false positive diagnosis in the context of degenerative diseases, the proposed system achieves sufficient recall rates for use as a primary diagnostic aid in medical settings lacking specialist expertise. These results indicate the significant potential of the proposed system as a diagnostic aid, particularly in medical settings lacking specialist expertise in interpreting bone radiographs, by providing a reliable initial assessment of obvious cases as well as a preliminary assessment of cases requiring more detailed specialist review.

The confusion matrix is a way to summarize the performance of a classification system. It not only provides information about the errors made by the classifier but also identifies the subtle errors that occur. The confusion matrix goes beyond simply relying on classification accuracy, as it also shows both correct and expected classifications, with the aim of highlighting areas for improvement and the strengths of the model. Figure 6 shows the confusion matrix of machine learning models with a selected feature set using HFHSA technology.

3.5.1 Discussion

The findings are categorical in that the combination of the Harmony search algorithm and machine learning methods has produced excellent classification performance of osteoporosis of the knee. The hybrid HFHSA method showed great performance enhancement of the traditional Harmony Search algorithm in most classification models and situations. XGBoost had an improvement of 0.92 percent (0.93589 to 0.941026) in classification, which is quite impressive in a clinical setting, where 1 percentage higher has the possibility of saving a life or preventing a misdiagnosis.

These findings verify that hybridization is also more successful in more complex tasks, where it is more efficient in feature selection and hyperparameter optimization resulting in better models that are more prepared to cope with various data complexities.

Also, when applying the feature selection methods, all the experimented ensemble algorithms Random Forest, Extra Trees, Gradient Boosting, Bagging, and XGBoost showed evident superiority to single Decision Trees. The underlying mechanisms of ensemble algorithms can be used to explain this superiority. These algorithms take the forecasts of multiple sub-models (trees) and minimize the variance of models by voting or averaging, and enhancing their capacity to predict new data. The data will contribute slightly different patterns to each individual tree and combining these patterns will result in reduced error caused by randomness and increased patterns that are not random.

3.5.2 Comparing the Results with Related Works

The results achieved in the research demonstrated the efficiency of the proposed system in classifying knee X-ray images with high accuracy compared to previous work where The proposed model achieved an exceptional accuracy of 94.87%, outperforming all previous studies mentioned. The comparison demonstrated that the proposed system was evaluated on a large dataset of 1,947 images, the largest dataset used in previous studies. This increases the reliability and generalizability of the results compared to some previous studies that relied on smaller datasets. It surpassed recent studies such as Dodamani and Danti (2023), who achieved 93.4% using DenseNet121 on 830 images, and Kumar et al. (2023), who achieved 93.5% using Inception v3 on only 240 images.

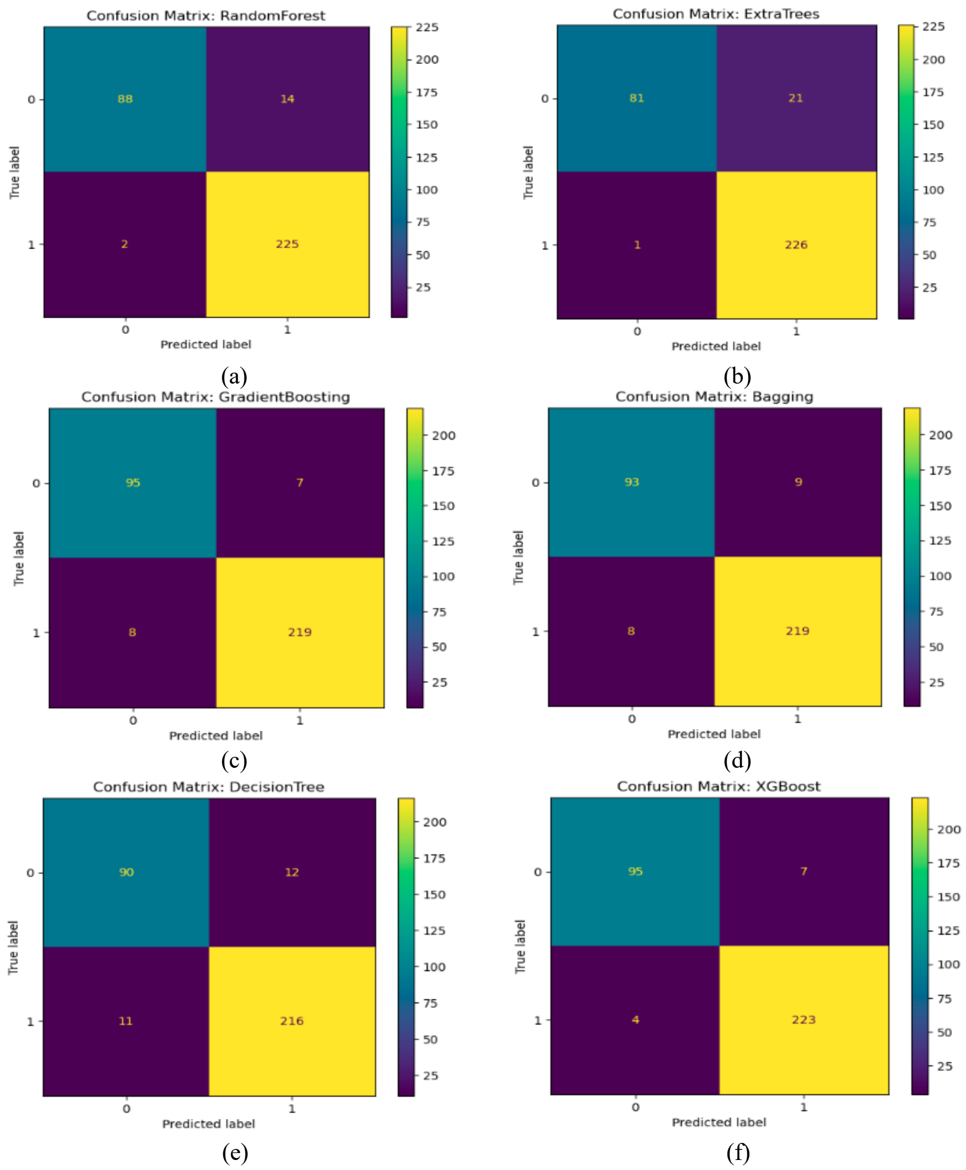

Figure 6: Confusion matrices for classification models using HFHSA: a) Random Forest; b) ExtraTrees; c) Gradient Boosting; d) Bagging; e) Decision Trees; f) XGBoost.

The clear superiority of the proposed system is due to several innovative technical factors, including the adoption of the HMMFF method to combine features extracted by traditional techniques with higher-level features such as YOLOv11, VGG-16, and ResNet-50, which helps provide a comprehensive and effective representation of bone features in X-ray images. The use of the hybrid feature selection technique (HFHSA) combines the power of the consensus search algorithm in global exploration with the flexibility of fuzzy logic in dealing with uncertainty, ensuring the selection of the best set of features and removing redundant features that may negatively impact system performance.

4 CONCLUSIONS

It has been noted that the detection of knee osteoporosis has gained importance over time due to its related complications especially with the emergence of artificial intelligence. The hybrid HFHSA and hybrid feature fusion (HMMFF) methods have made it possible to detect osteoporosis using X-ray images and these techniques have become the promising and effective technologies to enhance the accuracy and efficiency of detecting osteoporosis by the use of AI systems.

This research shows that integrating features obtained in various methods is valuable and efficient. Such integration provides an opportunity to use the diversity of patterns and structural information in images producing a complete and informative feature-vector. Moreover, it boosts the feature selection with the HFHSA technology that uses the efficient search of the Harmony Search algorithm and the capability of the GathGeva algorithm to deal with overlapping data points. This leads to the high classification accuracy which allows the accurate identification of the disease and classification of the affected area. The proposed system is an efficient tool to assist clinicians in making better and faster decisions and help to achieve better patient outcomes and lower treatment expenses. It should be pointed out though that these systems are not intended to take the place of the clinicians they are meant to aid them. The role of human clinical experience, medical judgment and capacity to synthesize various information about patients is also critical in the diagnostic and treatment process.

It is also recommended to continue developing osteoporosis detection techniques and to test the model on larger and more diverse datasets.

REFERENCES

[1] M. M. Sobh et al., "Secondary osteoporosis and metabolic bone diseases," J. Clin. Med., vol. 11, no. 9, p. 2382, 2022, [Online]. Available: https://doi.org/10.3390/jcm11092382.

[2] S. S. Amarnath, V. Kumar, and S. L. Das, "Classification of osteoporosis," Indian J. Orthop., vol. 57, no. Suppl 1, pp. 49-54, 2023, [Online]. Available: https://doi.org/10.1007/s43465-023-01058-3.

[3] M. Sarhan et al., "Knee osteoporosis diagnosis based on deep learning," Int. J. Comput. Intell. Syst., vol. 17, no. 1, p. 241, 2024, [Online]. Available: https://doi.org/10.1007/s44196-024-00615-4.

[4] I. M. Wani and S. Arora, "Osteoporosis diagnosis in knee X-rays by transfer learning based on convolution neural network," Multimed. Tools Appl., vol. 82, no. 9, pp. 14193-14217, 2023, [Online]. Available: https://doi.org/10.1007/s11042-022-13911-y.

[5] F. Rahim et al., "Machine learning algorithms for diagnosis of hip bone osteoporosis: a systematic review and meta-analysis study," Biomed. Eng. Online, vol. 22, no. 1, p. 68, 2023, [Online]. Available: https://doi.org/10.1186/s12938-023-01132-9.

[6] L. Gao, T. Jiao, Q. Feng, and W. Wang, "Application of artificial intelligence in diagnosis of osteoporosis using medical images: a systematic review and meta-analysis," Osteoporos. Int., vol. 32, no. 7, pp. 1279-1286, 2021, [Online]. Available: https://doi.org/10.1007/s00198-021-05887-6.

[7] S. Kumar, P. Goswami, and S. Batra, "Fuzzy rank-based ensemble model for accurate diagnosis of osteoporosis in knee radiographs," Int. J. Adv. Comput. Sci. Appl., vol. 14, no. 4, pp. 262-270, 2023, [Online]. Available: https://doi.org/10.14569/ijacsa.2023.0140430.

[8] S. Ponni, A. Sabarivani, and J. B. Janney, "Detection of Osteoporosis and Osteoarthritis Using Deep Learning Algorithms," in International Conference on Intelligent Vision and Computing, Springer, 2023, pp. 187-195, [Online]. Available: https://doi.org/10.1007/978-3-031-71391-0_15.

[9] P. S. Dodamani and A. Danti, "Transfer learning-based osteoporosis classification using simple radiographs," iJOE, vol. 19, no. 08, p. 67, 2023, [Online]. Available: https://doi.org/10.3991/ijoe.v19i08.39235.

[10] M. Shen, "Utilizing Deep Learning for Osteoporosis Diagnosis through Knee X-Ray Analysis," in 2024 International Conference on Artificial Intelligence and Communication (ICAIC 2024), Atlantis Press, 2024, pp. 553-560, [Online]. Available: https://doi.org/10.2991/978-94-6463-512-6_58.

[11] S. M. Naguib, M. K. Saleh, H. M. Hamza, K. M. Hosny, and M. A. Kassem, "A new superfluity deep learning model for detecting knee osteoporosis and osteopenia in X-ray images," Sci. Rep., vol. 14, no. 1, 2024, [Online]. Available: https://doi.org/10.1038/s41598-024-75549-0.

[12] A. OM and R. Gunasundari, "Advanced Osteoporosis Prediction from Knee X-rays via Residual Convolution and Recurrent Networks," Int. J. Intell. Eng. Syst., vol. 18, no. 6, 2025, [Online]. Available: https://doi.org/10.22266/ijies2025.0731.13.

[13] I. Muhammad, R. S. Rao, and B. Lee, "BONE-Net: A novel hybrid deep-learning model for effective osteoporosis detection," PLoS One, vol. 20, no. 10, p. e0334664, 2025, [Online]. Available: https://doi.org/10.1371/journal.pone.0334664.

[14] M. B. Qureshi et al., "Deep-learning based osteoporosis classification in knee X-rays using transfer-learning approach," Sci. Rep., vol. 15, no. 1, p. 38448, 2025, [Online]. Available: https://doi.org/10.1038/s41598-025-24338-4.

[15] Z. Salekshahrezaee, J. L. Leevy, and T. M. Khoshgoftaar, "The effect of feature extraction and data sampling on credit card fraud detection," J. Big Data, vol. 10, no. 1, p. 6, 2023, [Online]. Available: https://doi.org/10.1186/s40537-023-00684-w.

[16] E. Pashaei and E. Pashaei, "An efficient binary chimp optimization algorithm for feature selection in biomedical data classification," Neural Comput. Appl., vol. 34, no. 8, pp. 6427-6451, 2022, [Online]. Available: https://doi.org/10.1007/s00521-021-06775-0.

[17] R. Zubair and O. A. Alo, “Grey level co-occurrence matrix (GLCM) based second order statistics for image texture analysis,” arXiv Prepr. arXiv2403.04038, 2024, [Online]. Available: https://doi.org/10.20944/preprints202404.0881.v1.

[18] H. Shao, J. Pu, and J. Mu, “Pig-posture recognition based on computer vision: Dataset and exploration,” Animals, vol. 11, no. 5, p. 1295, 2021, [Online]. Available: https://doi.org/10.3390/ani11051295.

[19] M. Bansal, M. Kumar, and M. Kumar, “2D object recognition: a comparative analysis of SIFT, SURF and ORB feature descriptors,” Multimed. Tools Appl., vol. 80, no. 12, pp. 18839-18857, 2021, [Online]. Available: https://doi.org/10.1007/s11042-021-10646-0.

[20] L. S. Chow, G. S. Tang, M. I. Solihin, N. M. Gowdh, N. Ramli, and K. Rahmat, “Quantitative and qualitative analysis of 18 deep convolutional neural network (CNN) models with transfer learning to diagnose COVID-19 on chest X-ray (CXR) images,” SN Comput. Sci., vol. 4, no. 2, p. 141, 2023, [Online]. Available: https://doi.org/10.1007/s42979-022-01545-8.

[21] D. Theng and K. K. Bhoyar, “Feature selection techniques for machine learning: a survey of more than two decades of research,” Knowl. Inf. Syst., vol. 66, no. 3, pp. 1575-1637, 2024, [Online]. Available: https://doi.org/10.1007/s10115-023-02010-5.

[22] Fan, M. Chen, X. Wang, J. Wang, and B. Huang, “A review on data preprocessing techniques toward efficient and reliable knowledge discovery from building operational data,” Front. Energy Res., vol. 9, p. 652801, 2021, [Online]. Available: https://doi.org/10.3389/fenrg.2021.652801.

[23] F. Qin, A. M. Zain, and K.-Q. Zhou, “Harmony search algorithm and related variants: A systematic review,” Swarm Evol. Comput., vol. 74, p. 101126, 2022, [Online]. Available: https://doi.org/10.1016/j.swevo.2022.101126.

[24] M. Dubey, V. Kumar, M. Kaur, and T.-P. Dao, “A systematic review on harmony search algorithm: theory, literature, and applications,” Math. Probl. Eng., vol. 2021, no. 1, p. 5594267, 2021, [Online]. Available: https://doi.org/10.1155/2021/5594267.

[25] J. Wang, H. Ouyang, C. Zhang, S. Li, and J. Xiang, “A novel intelligent global harmony search algorithm based on improved search stability strategy,” Sci. Rep., vol. 13, no. 1, p. 7705, 2023, [Online]. Available: https://doi.org/10.1038/s41598-023-34736-1.

[26] Huang, “Feature Selection and Feature Stability Measurement Method for High-Dimensional Small Sample Data Based on Big Data Technology,” Comput. Intell. Neurosci., vol. 2021, no. 1, p. 3597051, 2021, [Online]. Available: https://doi.org/10.1155/2021/3597051.

[27] N. G. Peso’a and N. F. Gamayanti, “Implementation of the Gath-Geva clustering algorithm in the clustering districts/cities in Central Sulawesi based on public health development indicators,” in Proc. 4th Int. Seminar Sci. Technol. (ISST 2022), Atlantis Press, 2023, pp. 320-328, [Online]. Available: https://doi.org/10.2991/978-94-6463-228-6_36.

[28] X. Wu, H. Zhou, B. Wu, and T. Zhang, “A possibilistic fuzzy Gath-Geva clustering algorithm using the exponential distance,” Expert Syst. Appl., vol. 184, p. 115550, 2021, [Online]. Available: https://doi.org/10.1016/j.eswa.2021.115550.