International Journal of Computer Sciences and Engineering
Vol.**12**, Issue.**6**, pp.**01-12**, June **2024**
ISSN: 2347-2693 (Online)
Available online at: www.ijcseonline.org

Research Article

A Hybrid Model for denoising in Data Mining and Exploration

Cookey I.B.[1*], Bennett E.O.[2], Anireh V.I.E.[3], Matthias D.[4]

[1,2,3,4]Dept. of Computer Science, Faculty of Science, Rivers State University, Port-Harcourt, Nigeria

**Corresponding Author: cookey.ibiere@ust.edu.ng*

Received: 23/Apr/2024; **Accepted**: 25/May/2024; **Published**: 30/Jun/2024. **DOI**: https://doi.org/10.26438/ijcse/v12i6.112

Abstract: Image denoising is essential in digital image processing to improve the visual quality of images corrupted by noise. This study introduces a hybrid model combining fuzzy logic's adaptive capabilities with genetic algorithms' optimization power for effective denoising. The model leverages fuzzy logic to handle uncertainty and genetic algorithms to optimize denoising parameters. The hybrid model processes images in stages: a fuzzy logic-based filter preprocesses the noise-affected image, guided by fuzzy rules from a knowledge base. Concurrently, a genetic algorithm optimizes the filter's parameters through evolutionary techniques like crossover, mutation, and selection. The fuzzy logic and genetic algorithm components work together, with the fuzzy logic module using a Mamdani inference system and the genetic algorithm refining the denoised image Experimental results show the hybrid model outperforms traditional methods and standalone fuzzy logic or genetic algorithm approaches. Its adaptability allows it to handle varying noise levels and image content effectively, demonstrating robustness against different noise distributions. This makes it suitable for diverse denoising scenarios. The hybrid model represents a significant advancement in image denoising, highlighting the synergy between fuzzy logic and genetic algorithms. Future work will focus on further optimizations and extensions to improve applicability in real-world scenarios. Overall, this approach enhances noise reduction performance, image quality, and the preservation of important image features.

Keywords: Image Denoising, Fuzzy Logic, Genetic Algorithm, Datamining and Peak Signal to Noise Ratio.

1. Introduction

The swift advancements in data acquisition, transmission, and storage have precipitated an unprecedented expansion of extensive computerized datasets. These datasets are sourced from a myriad of sectors, including business, education, government, the scientific community, the Internet, and various offline and online repositories. They encompass diverse formats such as text, graphics, images, video, audio, animation, hyperlinks, and markups. Furthermore, these datasets are continually increasing in both the depth of their attributes and the breadth of their instance objects.

The advent of technology has significantly influenced the way people live their lives. During the early years, data computation was done manually, but today with the aid of computers data is easily generated, stored, and accessed.
The digital world tends to create data faster than ever. All data (information) is valuable. It is important to understand what data to collect, how to collect the data, and how to manage the data collected.

Various authors have reviewed the field extensively, despite the differing applications. According to Chau (2016), this field involves "the study of collecting, cleaning, processing, analysing, and gaining useful insights from data." Additionally, it is defined as "Data Science," which pertains to an emerging area focused on the collection, preparation, analysis, visualization, management, and preservation of large information collections Although the term Data Science appears to link most closely to these activities, it encompasses a broader scope that includes the collection, preparation, analysis, visualization, management, and preservation of extensive datasets.

Frenay and Kabán (2014) identified label noise as particularly detrimental to classification performance, as mislabelled examples hinder classification models from accurately learning the true distribution of the raw data. Consequently, developing an effective and reliable class noise detection method is essential for enhancing the noise adaptability of credit evaluation models. To address this noise issue, the use of a hybrid model (fuzzy logic and genetic algorithm) is introduced in this thesis.

Fuzzy logic is an approach to computing that operates on "degrees of truth" instead of the binary "true or false" (1 or 0) used in traditional Boolean logic. The term "fuzzy" refers to concepts that are unclear or ambiguous. As a branch of Artificial Intelligence, fuzzy logic can be utilized to develop intelligent systems that interpret information articulated in human language (Kambalimath et al., 2020). Fuzzy logic algorithms address problems by evaluating all available data to make the best possible decision based on the input, mimicking the human decision-making process that considers all possibilities between true and false. Fuzzy logic is an excellent tool.

In the fuzzy logic based denoising approach the intensity values of pixels in an image and it is represented by fuzzy sets. The fuzzy logic takes each pixel's neighbourhood and calculates the membership. Each pixel's intensity value is compared to threshold values defined for each fuzzy logic set. These threshold values define the range of intensity values for each fuzzy set. The membership function assigns a degree of membership to each fuzzy set based on how well the pixel's intensity value falls within the defined threshold range. However, the threshold values are based on the characteristics of the input image and the noise distribution.

The genetic algorithm (GA) draws its inspiration from Darwin's theory of biological evolution, embodying the principle of "survival of the fittest." In natural ecosystems, organisms that are better adapted tend to have a higher likelihood of survival (Mirjalili et al., 2019). The algorithm replicates this process through mechanisms such as individual replication, selection, inheritance, recombination, and variation of traits. Across successive generations, this process fosters the evolution of organisms in favorable directions, allowing for the accumulation of traits and the emergence of diversity within species. Consequently, populations undergo diversification, potentially leading to the creation of new species. Notably, genetic algorithms offer distinct advantages over other search algorithms, characterized by their simplicity of implementation, efficiency in target identification, and robustness in coding. As a result, they serve as potent stochastic search algorithms with widespread applications across diverse domains, including engineering, finance, robotics, and biology (Wang, 2022).

The genetic algorithm takes over and evolve the parameters of the fuzzy logic system. It generates the population of candidate solutions with each solution representing a set of fuzzy logic parameters. It then evaluates the fitness of each candidate solution using the fitness function which measures the denoising performance. The fitness candidate solutions are selected as parents for the next generation. Through successive generations, the genetic algorithm model applies crossover and mutation operations to explore the parameter space and improves the denoising performance. The optimized set of the fuzzy logic parameter that would be obtained after the optimization process represents the optimal configuration for denoising the salt and pepper noise in grey images.

2. Related Works

Mohammad-Rasool et al. (2020) conducted a study on predicting Iran's national grid's annual peak load and energy demand over ten years. They compared various forecasting techniques, including ARIMA, ANN, and Particle Swarm Optimization - Support Vector Regressor (PSO-SVR). The study found that ARIMA may lead to overestimation or underestimation, while PSO-SVR provided more accurate forecasts with small errors. The hybrid forecasting method, utilizing the circuit current division principle, was found to be powerful in leveraging the strengths of different methods. The study also highlighted concerns about potential power shortfall or planned load interruptions in Iran's grid, particularly during summer. The study suggests that clean renewable resources like wind and solar power should provide a significant portion of the required generation capacity.

In brief, Ougiaroglou et al. (2015) investigated two approaches for managing noisy data in K-NN classification: algorithmic editing and the utilization of a high K value. The research conducted revealed that employing a larger K value resulted in increased accuracy, albeit necessitating expensive iterative adjustments. In contrast, editing algorithms offered a more straightforward methodology albeit with marginally diminished precision. The report proposed potential avenues for further research, such as the creation of non-parametric adaptive classification models that may dynamically modify the number of nearest neighbors according to the quality of the data.

Puri's 2022 paper presents an enhanced hybrid bag-boost model for handling noisy class imbalance datasets. The model uses the K-Means SMOTE-ENN resampling technique, which is robust in handling increasing noise levels among imbalanced classes. The study evaluates benchmark binary class imbalanced datasets and finds the hybrid bag-boost technique efficient compared to bagging and AdaBoost using decision tree learning techniques. The model can be extended to multiclass noisy imbalanced datasets and other applications but may face computational cost constraints.

Puri (2022) presents a hybrid bag-boost approach for managing noisy class imbalance datasets, using the K-Means-SMOTE–ENN resampling technique. The model improves performance on binary class imbalanced datasets, regardless of noise. Further investigation is recommended for multiclass noisy imbalanced datasets.

3. Methodology

Although there are several research methods suitable for the carrying out of this study, for the proposed study, a Constructive Research Methodology is adopted, while the Object-Oriented Design Approach (OODA) is adopted for the system structural development.

Dataset Description
The data used, are salt and pepper noisy grey images.

System Design
"In the development of the hybrid model, components will be conceptualized as objects, and the OODA (Observe, Orient, Decide, Act) framework will be employed in its design. This novel construct will be evaluated using both hard and soft paradigms to rigorously demonstrate its functionality, owing

to its integration of both theoretical and practical dimensions. The efficacy of the model will be validated through cross-validation techniques."

The proposed system is a hybrid model that integrates fuzzy logic model and genetic algorithm model. Each of these models has a specific role in handling noisy image data.
The initial step in preprocessing the noisy image involves the utilization of a fuzzy logic-based adaptive filter. This filter employs the fuzzy inference system (FIS) to assess the level of noise present in each pixel and subsequently perform the necessary denoising operation. A fuzzy filter is utilized to initially reduce noise in the noisy image. Fuzzy logic is utilized to evaluate the level of noise present in individual pixels by considering the intensity and values of the surrounding pixels. The classification of pixels as noisy or noise-free is determined by fuzzy rules, which are established based on their membership to different linguist variables that reflect different levels of noise, such as low, high, or medium. After employing fuzzy inference to ascertain the noise level of individual pixels, denoising methods are subsequently utilized to approximate the actual intensity of the noisy pixels. These guidelines are formulated using expertise in the field and practical observations, enabling precise denoising while maintaining the integrity of visual information.

The adaptive filter, which is based on fuzzy logic, adjusts its denoising technique by considering the attributes of the input image and the observed distribution of noise. The inherent adaptability of the filter enables it to proficiently manage diverse noise patterns and intensities found in practical situations.

The genetic algorithm model initializes the output from the fuzzy logic model. This model would be used to optimize the membership function to give optimal solutions. Genetic Algorithm model is used as a tuning factor to optimize the fuzzy rules generated by the fuzzy logic model. Genetic algorithms excel in global optimization and search spaces. When used in conjunction with fuzzy logic, genetic algorithms can explore a wide range of parameter configurations to find an optimal or near-optimal solution for denoising, considering the complexity of image data. The Genetic algorithms model is employed to optimize the parameters of the fuzzy logic system for denoising. This includes tuning membership functions, rule strengths, and other parameters to enhance the performance of the denoising algorithm. It will contribute to the robustness of the hybrid model by avoiding local minima and finding solutions that generalize well to different images. This is crucial for creating denoising models that work effectively across various types of images and noise patterns. It will leverage parallel processing, which is highly beneficial for speeding up the optimization process. In the context of image denoising, this will lead to faster and more efficient convergence towards optimal denoising parameters. Both fuzzy logic and genetic algorithms are effective in handling non-linear relationships. Salt and pepper noise introduces a non-linear distortion in images, and the hybrid model can better capture and mitigate these distortions.

Combining fuzzy logic and genetic algorithms, the hybrid model leverages the fuzzy system's ability to handle uncertainty and complex relationships with the global optimization capabilities of genetic algorithms. This synergy results in a robust and adaptive approach for effectively handling salt and pepper noise in digital images.

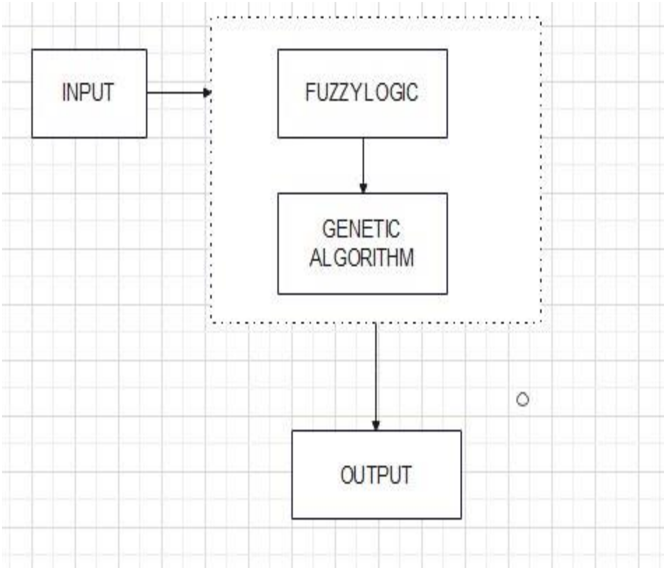

Figure 1. Schematic view of the proposed system architecture.

Figure 1 captures the schematic architectural design of the proposed hybrid model. It depicts the flow of the denoising process However, this study is restricted to developing a hybrid model in denoising images. Hence, it will be limited to denoising salt and pepper noise in grey images.
The proposed system architecture is captured in Figure 2.

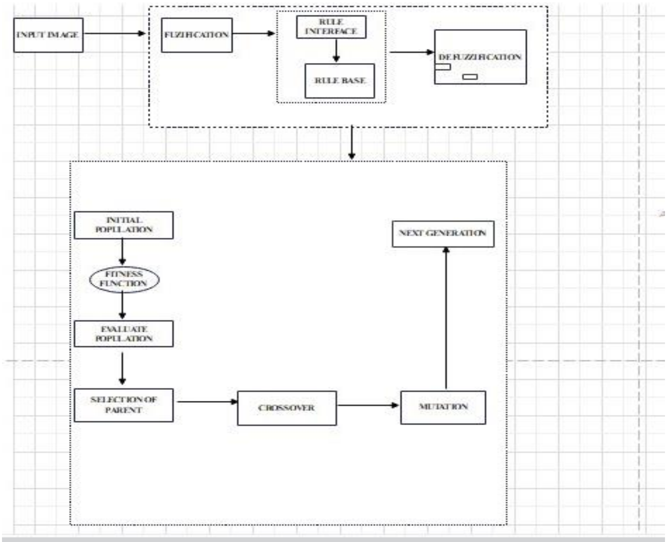

Figure 2. Proposed System Architecture.

Figure 2. represents a high-level architectural design of the system for denoising salt and pepper grey images. It captures the integration of Fuzzy logic model and Genetic Algorithm model.
The dataset consists of a noisy image obtained from the internet.

An adaptive filter based on fuzzy logic. This is a method of filtering the mean. This filter is based on fuzzy logic is a modified version of the median filter that integrates the efficiency of the mean filter with the resilience of the median filter. The fundamental concept involves the iterative application of a mean filter until the resultant image reaches a level of convergence towards stability. The filter can be represented as an iterative procedure: Let I (x, y) denote the pixel value located at coordinates (x, y) in the input image, and O (x, y) represent the corresponding pixel value in the output picture subsequent to the application of the Fuzzy logic-based adaptive filter. Let I(x, y) be the pixel value at coordinates (x, y) in the input image, and O (x,y) be the corresponding pixel value in the output image after applying the FastNMean filter.

$$0(x, y, k+1) = {}^{1}/_{W} \sum i\hat{}W \,[=1]\, I(xi, yik) \quad (1)$$

Where: O (x, y, k, +1) is the pixel value at coordinates (x, y) in the output image after k + 1iterations.
W is the size of the filter window.
(xi, yi) are the coordinates within the filter window, centered at (x, y). I (xi, yi, k) is the pixel vale at coordinates (xi, yi) in the input image after k iterations.
The iteration process would continue until convergence, and the resulting image O (x, y) is considered the denoised output.

Data Preprocess
A crucial step in the hybrid paradigm for managing noise in multimedia mining is data preprocessing. The process involves a series of procedures that are executed on the input data in order to ready it for the subsequent phases of feature extraction and noise reduction. The preprocessing procedures are employed to preprocess the signal, threshold values and noise level.

The initial phase of the hybrid model is of utmost importance in guaranteeing that the input data is adequately prepared for the subsequent stages of feature extraction and noise removal. Through the implementation of many techniques such as data cleaning, normalization, resampling, feature extraction, feature selection, and data augmentation, the hybrid model demonstrates a high level of efficacy in the identification and elimination of noise from multimedia data that is inherently noisy. Consequently, this leads to enhanced data quality and accuracy.

In this step, the image is collected and preprocessed to remove any noise that may be present in the data. This can involve tasks such as data cleaning, filtering, and feature extraction.

Let X be the input noisy data, and Y be the output denoised data. Let f() be a denoising function that takes X as input and outputs Y. Then the mathematical expression for data preprocessing noisy data can be written as:

$$Y = f(X) \quad (2)$$

In this context, the denoising function f () is employed to perform a series of operations on the noisy input X, aiming to diminish the noise and produce a more refined output Y. These operations encompass many statistical approaches such as filtering, smoothing, and others, which are employed to detect and eliminate noise from the input data.

The denoised data Y from the output can then be utilized for additional processing and analysis in multimedia mining applications, including but not limited to picture or video categorization, object detection, and content-based retrieval.
Fuzzy Logic Model:

The purpose of this Fuzzy Logic model is to address uncertainty and imprecision by integrating degrees of truth instead of binary values (true or false). The thesis employs the Mamdani inference approach. Fuzzy logic functions utilize membership functions to allocate degrees of membership to components inside a set, considering the intensity of the pixels (low, medium, and high). These functions quantify the unpredictability in data by depicting the gradual shift between being a member and not being a member. The sigmoidal membership functions are which is expressed as:

$$\mu(x) = \frac{1}{(1+e^{-\alpha(x-\beta)}))} \quad (3)$$

where x is crips input value
α is the parameter that controls the slope of the sigmoid curve.
β is the parameter that determines the position of the sigmoid midpoint
e is the base of the natural logarithm.

The sigmoid function produces an S-shaped curve, and the parameters α and β would allow the shape and position of the curve to be adjusted. The larger the value of α, the steeper the slope of the curve, resulting in a more abrupt transition between low and high membership values. The parameter of β shifts the midpoint of the curve along the x-axis.
Fuzzification: Inputs are transformed into fuzzy values using appropriate membership functions.

Rule Inference: here the fuzzy rules are applied to make decisions or predictions based on fuzzy inputs. The rule applied and the decision is fed into the Rule base.
Rule Base uses rules that linguistically describe relationships between input and output variables. These rules would be expressed in the form of "if-then" statements using linguistic terms (e.g., "if the pixel intensity is high, then decrease pressure slightly") and then output as fuzzy set which would be input to be defuzzified, which is also known as defuzzification.

The input is converted into fuzzy steps and then the matching degree is compared between the input and the rules. After matching degree, the system determines which rule to be added based on the input field. When all the rules are fired, the fired rules would combine to develop the action control. Finally, the generated inputs are transformed into crisp value.

Defuzzification The fuzzy set (output) is converted back into a crisp value for practical application or further analysis.
Mathematical representation of the fuzzy logic denoising model.

Let I (x, y) be the pixel value/intensity of the noise image at coordinates (x, y), let μ A (I (x, y)) represent the membership degree of the pixel intensity/value I (x, y) in the fuzzy set A. the fuzzification equation could be linear or nonlinear mapping.
μA (I (x, y)) =fuzzification function (weighted mean) (I (x, y))
An example using linear mapping for μA (I (x, y)) =fuzzification function (weighted mean) (I (x, y)).

$$\mu A\left(I\,(x,y)\right) = \frac{I(x,y) - \mathrm{Min}_{\mathrm{Intensity}}}{Max_{Intensity} - Min_{Intensity}} \qquad (4)$$

where:
Min_Intensity is the minimum possible intensity value in the image.
Max_Intensity is the maximum possible intensity value in the image.
The linear mapping scales the intensity values to the range of [0,1] indicating the degree of membership in the fuzzy set.
Fuzzification with weighted mean

$$\mu A\,(I\,(x,y)) = \frac{\sum_{i,j} w(i,j).I(x+i,y+j)}{\sum_{i,j} w(i,j)} \qquad (5)$$

where:
w (i, j) is the weight assigned to the pixel at the relative coordinates (i, j) in the local neighborhood.
Note that the weight w (i, j) can be determined based on the proximity of the pixel (x + I, y +j) to the target pixel (x, y)
Fuzzy Inference here the fuzzy logic rules are defined to infer denoised intensity values on the fuzzified intensities.
Denoised Intensity (x, y) = fuzzyinference

$$(\mu A(I(x,y))) \qquad (6)$$

Given a scenario where we have an original image (5 ×5 pixels, intensity values between 0 and 255):

100	150	200	50	75
125	180	40	210	30
90	160	70	220	110
30	200	100	120	180
80	110	140	60	190

SALT AND PEPPER NOISY IMAGE MATRIX.
Let's assume that 10% of the pixels in the original image are affected by salt and pepper noise, randomly set to either 0 or 255.
Defining the Fuzzy sets and membership functions: we define three fuzzy sets to "Low", "Medium" and "High". (we would use sigmoidal membership function)
Where I(x, y) is the represents the intensity value of the pixel at coordinate (x,y) in the image.

μLow(z)=1/(1+ e^(-0.1(z-100)))
μMedium(z)=1/(1+ e^(-0.1(z-150)))
μHigh(z)=1/(1+ e^(-0.1(z-200)))
given e as the natural logarithm.
The fuzzy rules would be:
Rule 1: if Low and the neighborhood is Medium, then reduce noise influence.
Rule 2: If Bright, then prioritize the denoised value.
Then, the fuzzy inference is applied to each pixel considering its membership in the fuzzy sets and the rules are applied.

Denoised intensity (x,y)

$$= \frac{\mu Dark\,(I(x,y)).intensity_{Dark} + \mu Medium\,(I(x,y)).intensity_{Medium} + \mu Bright\,(I(x,y)).intensity_{Bright}}{\mu Dark(I(x,y)) + \mu Medium(I(x,y)) + \mu Bright(I(x,y))}$$

Assuming Intensity Dark =50
Intensity Medium = 150
Intensity Bright =200
I(x, y) = I(3×4) = 220
Substituting the values into the equation:

Denoised intensity (220) =

$$\frac{\mu Dark\,(I(220)).50 + \mu Medium\,(I(220)).150 + \mu Bright\,(I(220)).200}{\mu Dark(I(220)) + \mu Medium(I(20)) + \mu Bright(I(220))}$$

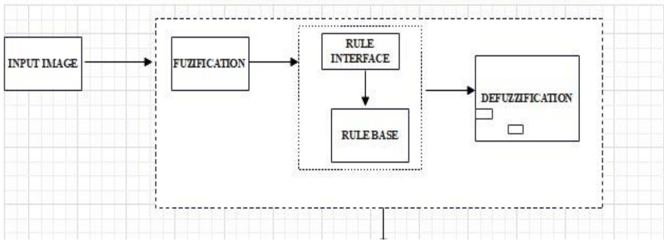

Figure 3: Fuzzy Logic Module

Genetic Algorithm Model
Genetic Algorithm: The genetic algorithm component of the hybrid model is used to optimize the set of rules created by the fuzzy logic component. The genetic algorithm can be used to fine-tune the fuzzy rules to better handle noisy data (i.e. being a tuning factor).

Genetic algorithms, inspired by natural evolution, would provide a powerful optimization technique for identifying the optimal set of fuzzy rules. By mimicking the processes of natural selection, mutation, and recombination, genetic algorithms iteratively refine the population of fuzzy rules, gradually improving their ability to accurately detect and suppress noise in the signal. This evolutionary approach allows the system to explore a vast search space of potential rule combinations, effectively avoiding the limitations of traditional optimization methods that may become trapped in local optima.

The genetic algorithm component generates a set of fuzzy rules that will be used by the fuzzy logic module to determine the degree of noise present in the image. Genetic algorithms, on the other hand, are useful in optimizing complex systems through the process of evolution, mimicking natural selection.
The incorporation of the genetic algorithm stage is vital in this hybrid model's ability to manage noise effectively.

Genetic algorithms are optimization algorithms inspired by natural selection and genetics. They work by evolving a population of potential solutions over generations, with each generation improving on the previous one through a process of selection, crossover, and mutation. The genetic algorithm stage in this hybrid model for handling noise in multimedia data typically involves the following steps:

1. Fitness Function involves scoring each potential solution based on how well it performs in removal of noise from the input data (image data), due to the high dimensionality, heterogeneity and variability of the multimedia data. By incorporating these considerations, the fitness function provides a comprehensive evaluation of potential solutions in denoising multimedia data. It captures the high dimensionality, heterogeneity, and variability of the data, ensuring that the selected fuzzy rule sets or solutions are capable of effectively removing noise across the image data, noise patterns, and application scenarios. The genetic algorithm component utilizes this fitness evaluation to guide the evolution and optimization process, ultimately leading to the selection of fuzzy rule sets that demonstrate superior performance in denoising image.
2. Selection involves choosing the best-performing solutions from the current generation to be used as parents for the next generation. The selection process is typically based on fitness, with better-performing solutions having a higher probability of being selected. By iteratively selecting the best-performing solutions and using them as parents for the next generation, the genetic algorithm effectively mimics the process of natural selection, leading to the gradual evolution of fuzzy rules that are better suited for noise reduction in image data. This evolutionary approach would enable the system to adapt to the complex and variable nature of the image data, achieving effective noise reduction without compromising the quality of the image.
3. Crossover involves combining the selected solutions to create new solutions for the next generation. This would be done by randomly selecting a subset of the parameters from each parent solution and combining them to create a new solution.
4. Mutation: Mutation involves randomly modifying a small subset of the parameters in the new solutions to introduce diversity and prevent the algorithm from getting stuck in local optima.

The steps of fitness evaluation, selection, crossover, and mutation are repeated for multiple generations until a satisfactory solution is reached.
The genetic algorithm stage of the hybrid model for handling noise in data plays an important role in optimizing the noise removal process. By using a population-based approach and evolving potential solutions over multiple generations, the genetic algorithm can effectively search for optimal parameter settings to improve noise removal performance. The combination of fuzzy logic and genetic algorithms in the hybrid model allows for a powerful and flexible approach to handling noise in multimedia data.
Mathematical representation of the genetic algorithm model:
Chromosome Representation,

Each gene in the chromosome corresponds to the intensity value of a pixel.
That is: chromosome = $[G_1, G_2, \ldots\ldots..G_n]$

Where Gi is the intensity value of the i^{th} pixel in the image.
Fitness function would determine how well the denoised image would match the original image and the choice of the pixel wise difference metric is the Mean Squared Error (MSE), hence lower values indicate better similarity.

Mean Squared Error is a widely used metric for quantifying the difference between two images, typically an original/reference image and a processed/modified image. It's a measure of the average squared differences between corresponding pixels in the two images.

Here's how it's calculated:
First step, the image difference is obtained by Subtracting each pixel value in the processed image from the corresponding pixel value in the original image. This results in a new image where each pixel represents the difference between the two images. Secondly, the difference is squared, square each difference value obtained in the previous step. This ensures that all differences are positive, emphasizing larger discrepancies. Third step Compute the Mean: Calculate the average of all the squared differences. This is done by summing up all the squared differences and dividing by the total number of pixels in the image.

Thus: fitness (Chromosome) = some measure of similarity with the original image.
= MSE (Original Image, Denoised Image)
Selection Probability (Chromosome)

$$=\frac{fitness\ (chromosome)}{\sum fitness\ (chromos0me)} \quad (8)$$

Crossover combines the genetic material from two parents by randomly selecting genes from each parent with equal probability:
Crossover (Parent1, Parent 2) = combined genes from Parent 1 and parent 2
That is:

$$childi = Random(Parent_1[i], Parent2[i] \quad (9)$$

Mutation random changes in individual chromosomes to explore new regions of the solution space is introduced. The intensity values of the pixels in the chromosome remains within the intensity range of [0, 255, 1] for the 8-bit grayscale image.

Mutation (chromosome) = Randomly alter some genes in the chromosome

Mutated chromosome [i] = chromosome [i] + Random Noise (10)

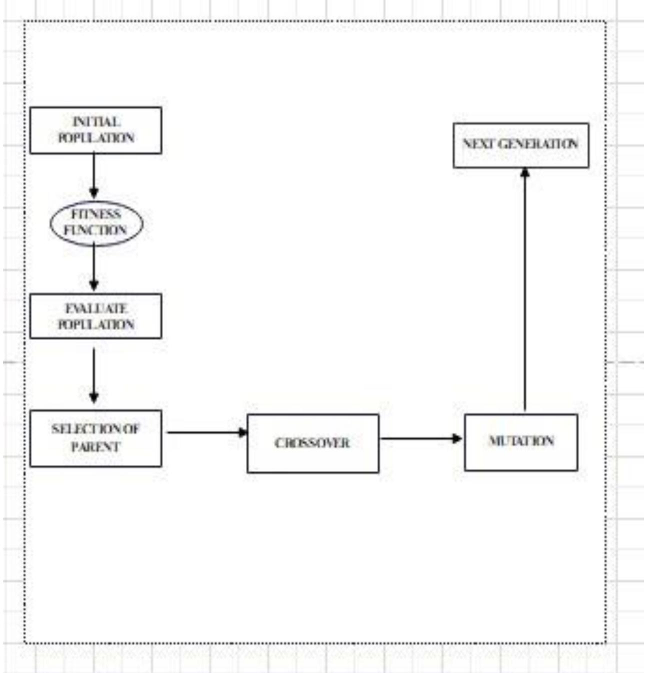

Figure 4: Genetic Algorithm Module

Table 1. Input Specification Design.

Variables/Parameters	Description
Input data (multimedia data)	The data that will be analyzed for noise
Fuzzy sets	The utilization of fuzzy sets to depict the diverse degrees of noise contained in the input data.
Fuzzy rules	Based on the fuzzy sets, a series of rules will be employed to deduce the intensity of noise present in the input data.
Genetic algorithm parameters	The parameters that will be used to tune the genetic algorithm, such as the population size, mutation rate, and crossover rate.
Integration strategy	The combination of the fuzzy logic system and the genetic algorithm involves utilizing the fuzzy logic system to develop initial candidate solutions for the genetic algorithm.
Output	The hybrid model's output, consisting of PSNR values obtained by calculating the number of generations required to achieve all objectives, typically aims for a PSNR value exceeding 30 dB. This threshold is commonly used as a starting point for denoising activities to ensure quality, computing efficiency, and other relevant criteria. The noise originating from the input data can be utilized.

Hybrid Model:

The hybrid model combines the strengths of fuzzy logic in handling uncertainty with the optimization capabilities of genetic algorithms to manage noisy data.

Genetic Algorithm with fuzzy logic makes up the hybrid model. Fuzzy logic's capacity to deal with ambiguity and imprecision is combined with genetic algorithms' capacity for optimization in this hybrid model. Fuzzy logic, with its linguistic concepts and membership functions, accommodates the inherent uncertainty in noisy data, making this synergy particularly useful for processing noisy data. The fuzzy logic model performs better in the presence of noise when the parameters are optimized, and the right features are chosen using genetic algorithms.

When working with data that is noisy, uncertain, or imprecise, the hybrid approach makes the most of the advantages of both techniques to produce a model that is more reliable and accurate.

Hybridization of Fuzzy Logic and Genetic Algorithms:

Genetic algorithms evolve the fuzzy logic system by refining the fuzzy rules or parameters based on their performance in noisy environments.

Noise Tolerance and Robustness:

The hybrid model's adaptive nature helps in building models that are more tolerant to noise and capable of making accurate decisions even with noisy inputs.

Mathematical Representation of the hybrid model

Hybrid Model = Genetic Algorithm (Fuzzified DenoisedImage)

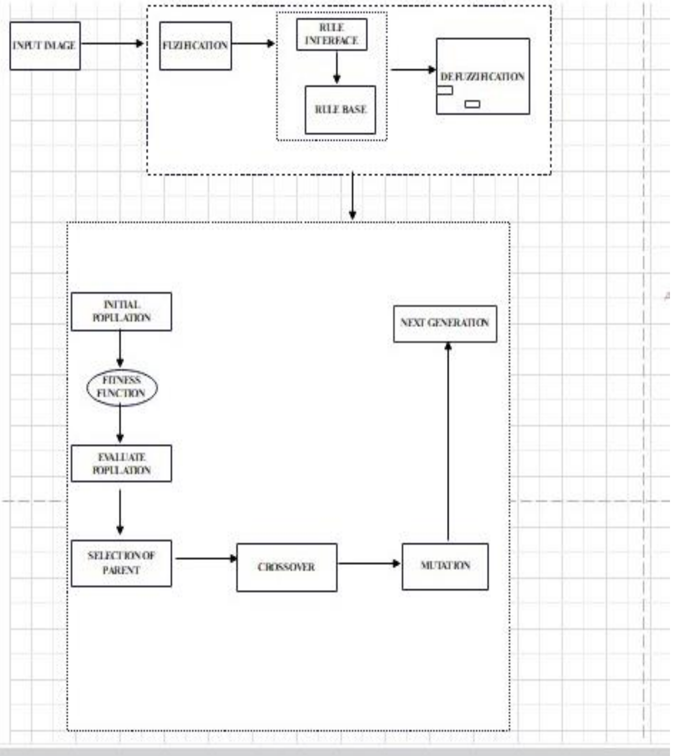

Figure 5: Hybrid Model

The output of the hybrid model is a set of weights or coefficients that are used to filter the noise from the input data. These weights are applied to a filtering algorithm or a signal processing technique to remove the noise from the input data.

Noise Detection Result: The primary output of the hybrid model is a binary result that indicates whether the input data contains noise or not. This result is generated by the fuzzy logic system and is optimized using the genetic algorithm.

Noise Detection Performance Metrics: In addition to the noise detection result, the hybrid model may also output performance metrics (PSNR) that evaluate the performance of the system. Some commonly used performance metrics include accuracy, precision, recall, F1 score, and area under the receiver operating characteristic curve (AUC-ROC).

Fuzzy Logic System Parameters: The hybrid model outputs the parameters of the optimized fuzzy logic system. This can be useful for further analysis and for understanding how the system makes decisions.

Genetic Algorithm Optimization Details: The hybrid model may output details about the genetic algorithm optimization process, such as the best solution found, the convergence curve, and the number of generations required to converge.

Overall, the output of a hybrid fuzzy logic and genetic algorithm model for detecting noise in grey image provides important information about the presence and magnitude of noise, as well as potential solutions for reducing or removing that noise and improving the quality of the data.

Output: The output of a hybrid (fuzzy logic and genetic algorithm) model for detecting noise in grey image typically includes several variables that provide information about the noise level and characteristics of the input image. The specific output variables and their values depend on the particular application and context of the model, as well as the specific input variables and sources.

One common output variable is the noise level, which indicates the degree or magnitude of noise present in the input image. This can be represented by a numerical value or a visual representation, such as a color map or heat map. Another output variable is the noise type, which may indicate the specific type of noise present in the image for this paper the type of noise is the salt and pepper noise. This noise (salt and pepper) value ranges from 0 (black) and 255 (white) in 8-bits grayscale image. In denoising there are filter coefficient which are also known as filter kernel, which are finite number of bits, commonly in the range of 8 to 16 bits. The specific values of the filter coefficients depend on the type of filter and the desired filtering operation Regarding this paper, the adaptive media filter technique (it dynamically adjusts the filter coefficients to achieve both noise reduction and detail preservation) note that performance of the median filter depends on the size of the filter window, which determines the number of neighboring pixels used to calculate the median value. For salt and pepper noise, a smaller filter window would be preferred, as it helps preserve image details while effectively removing noise. However, using too small a filter window can lead to over-smoothing and loss of fine image features. The optimal filter window size depends on the characteristics of the noise and the specific image being processed.

The hybrid model may also generate filter coefficients that can be applied to the input data to reduce or remove noise. These filter coefficients can be represented as a set of numerical values or as a mathematical function.
In addition to noise-related variables, the hybrid model may also output quality scores for the video and audio components of the input image. These quality scores indicate the overall perceived quality of the denoised grey image and can be used to evaluate the effectiveness of the noise detection and reduction process.

4. Results and Discussion

This interface enables the user select salt and pepper noisy grey image, denoise the image and visualize the image. The Peak signal to noise ratio of 10 generations is calculated, the average of each generation PSNR is displayed at the end of every generation calculated. Solutions from each generation is applied to the salt and pepper noisy image.

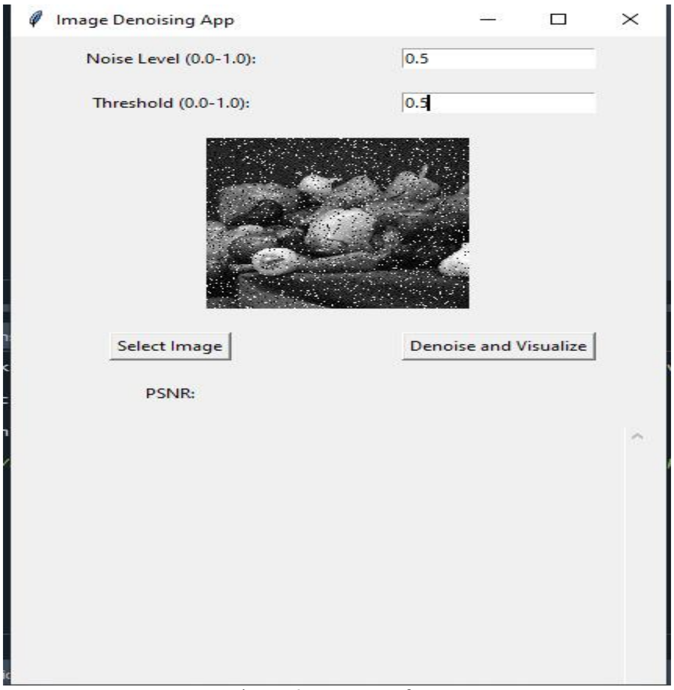

Figure 6: Input Interface.

This interface shows the sample of the salt and pepper image, which the hybrid model tends to denoise.
After each generation, the fitness of each individual which is a combination of the noise level and the threshold parameters is evaluated of each population. The fitness function is used to measure the quality of the denoised image. In this case the fitness function calculates the Peak Signal-to-Noise Rate (PSNR) between the original noisy image and the denoised image. After the fitness of each individual in the population is evaluated, the fitness value is stored in an array. This array contains the fitness values of all individuals in the current generation.

The average fitness is calculated when the array is populated, it calculates the average fitness of the current generation by summing up all the fitness values in the array and divides the total by the number of individuals which is also known as the solutions in the population.

This average fitness gives an indication of how well the population as a whole performs in each generation of the genetic algorithm. This is an essential metric for tracking the progress of the optimization process.

For the first generation, the average fitness 44.47261651583694 simply indicates that the average fitness of the individuals in the population is approximately 44.47.

The sizes-fitness (10), population (10, 2) gives information of the size of the fitness array and the population for the first generation. (10) Indicates that the fitness array has 10 elements, corresponding to the fitness values of the 10 individuals in the population. (10, 2) this simply means that the population has 10 individuals, and each individual has 2 parameters (noise level and threshold) size of offspring: (6, 2) is the size of the offspring generated through crossover and mutation operations. The 6 is the offspring individuals and each individual has 2 parameters. Size fitness (10), Population (10, 2), New Population (12, 2) this provides the information about the size of the population after combining the parents and offspring. (12, 2) is the new population size of 12 individuals with each individuals having 2 parameters. The population size increased from 10 to 12, due to the addition of the offspring generated through the crossover and mutation.
This would be seen through the ten generations.

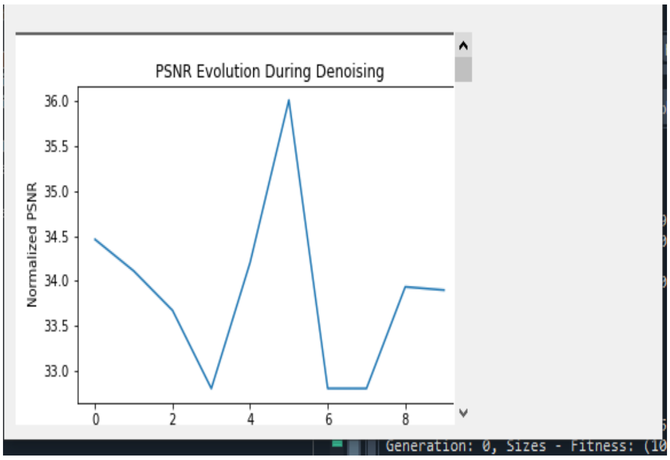

Figure 7

SE
Second Generation, the average fitness 41.889240550999954 simply indicates that the average fitness of the individuals in the population is approximately 41.88.

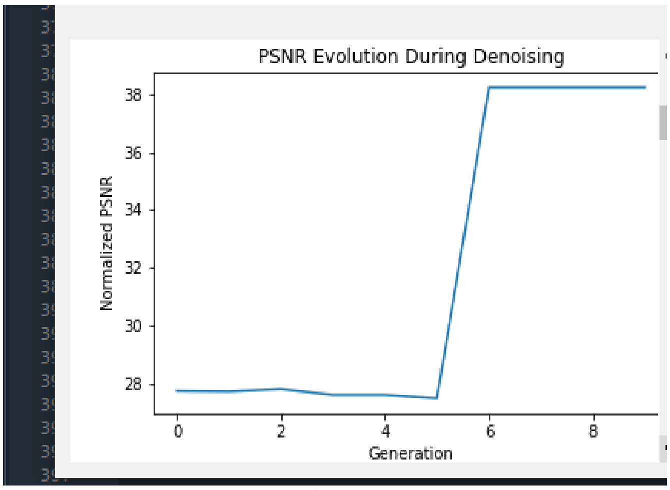

Figure 8

Third Generation, the average fitness 51.94903645055456simply indicates that the average fitness of the individuals in the population is approximately 51.94.

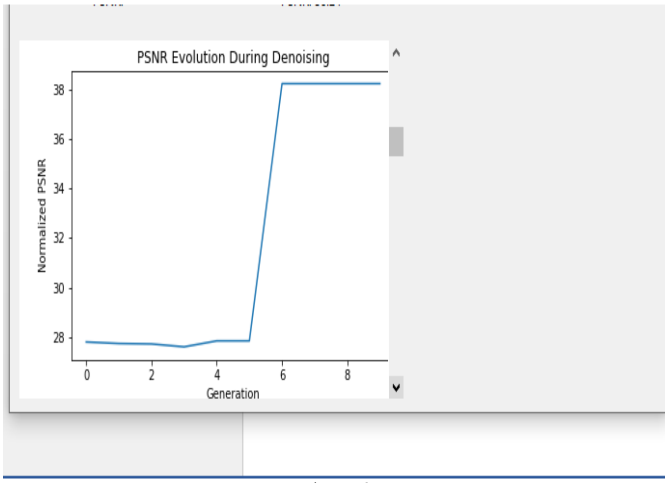

Figure 9

Fourth Generation, the average fitness 41.882034043110888 simply indicates that the average fitness of the individuals in the population is approximately 41.88.

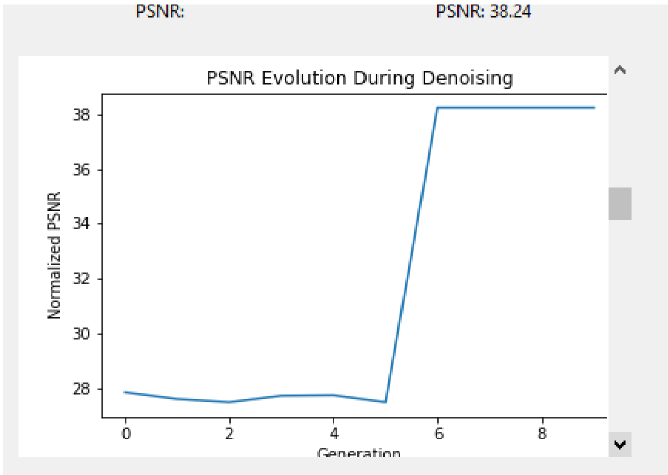

Figure 10

Fifth Generation, the average fitness 41.939422177542077simply indicates that the average fitness of the individuals in the population is approximately 41.93.

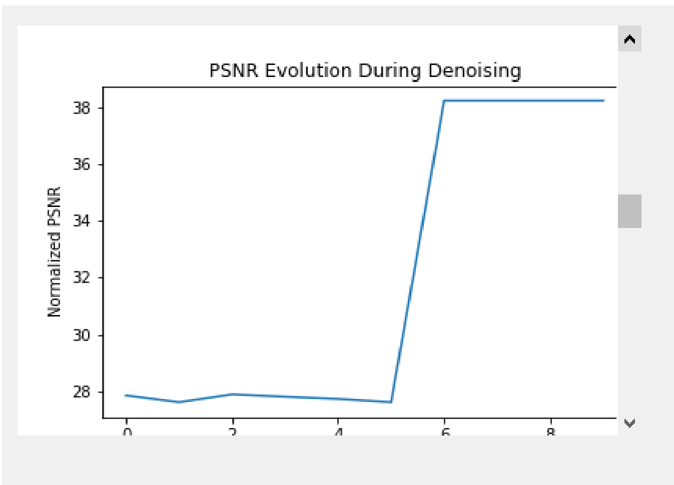

Figure 11

Sixth Generation, the average fitness 41.959075459052535 simply indicates that the average fitness of the individuals in the population is approximately 41.95.

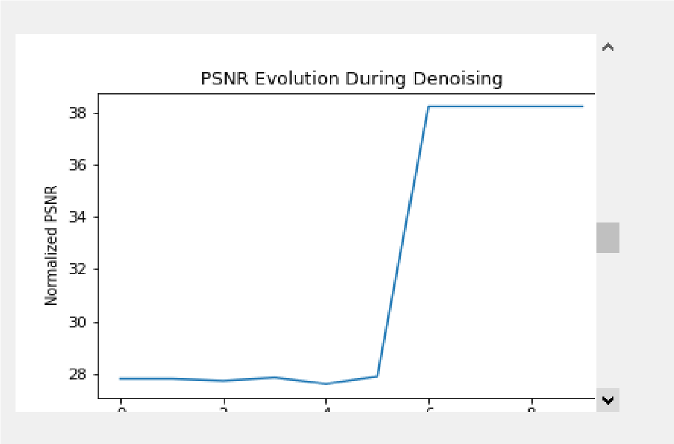

Figure 12

Sixth Generation, the average fitness 59.862978928659782simply indicates that the average fitness of the individuals in the population is approximately 59.86.

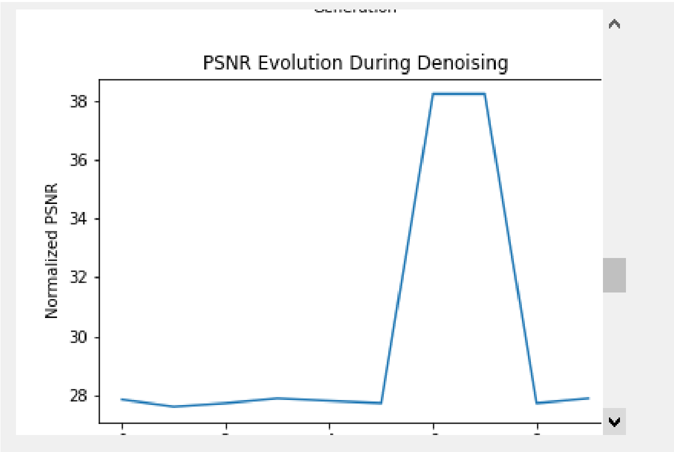

Figure 13

Seventh Generation, the average fitness 40.349256196241793 simply indicates that the average fitness of the individuals in the population is approximately 40.34.

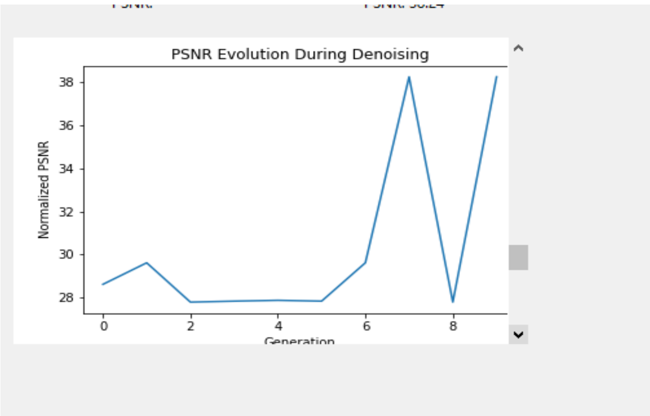

Figure 14

Eighth Generation, the average fitness 40.965911654497575 simply indicates that the average fitness of the individuals in the population is approximately 40.96.

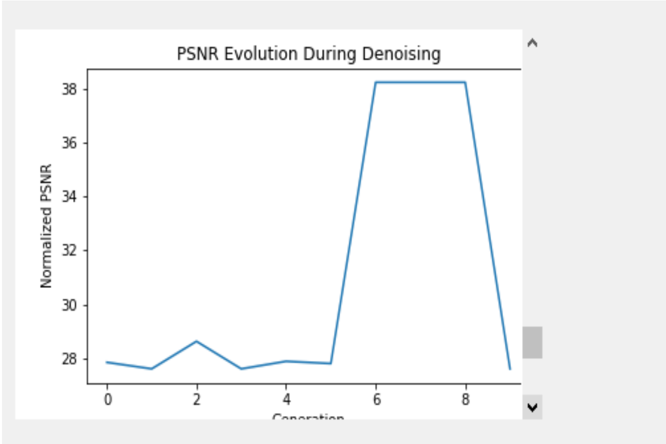

Figure 15

Nineth Generation, the average fitness 52.941438572616725simply indicates that the average fitness of the individuals in the population is approximately 52.94.

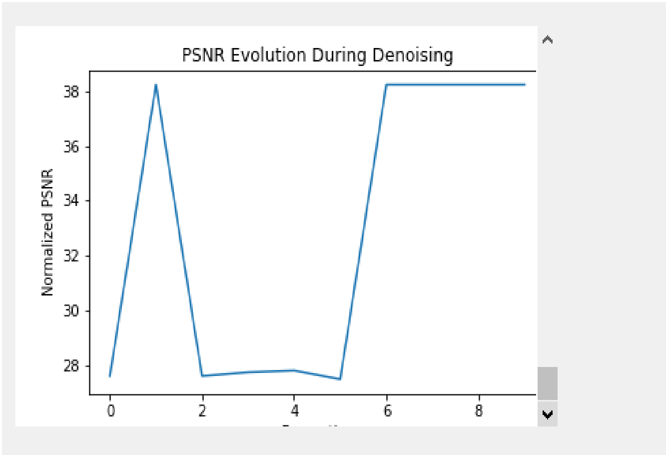

Figure 16

Results Summary

The results of our study are tabulated in Table 1. The table captured the Number of generations, Average Fitness (PSNR), Size of Offspring and the Fitness size This dataset contains salt and pepper grey noisy images. The results of the developed system when implemented is captured in Table 2.

Table 1. Results

Number of Generations	Average Fitness (PSNR)	Size of Offspring	Fitness Size
0	44.47	6, 2	10
1	41.88	6, 2	10
2	51.98	6, 2	10
3	41.88	6, 2	10
4	51.93	6, 2	10
5	41.95	6, 2	10
6	59.86	6, 2	10
7	40.34	6, 2	10
8	50.96	6, 2	10
9	52.94	6, 2	10

Table 2. Evaluation Performance Matrix (PSNR)

Number of Generations	Average Fitness (PSNR)	Quality	Size of Offspring	Fitness Size
0	44.47	Excellent	6, 2	10
1	41.88	Excellent	6, 2	10
2	51.98	Excellent	6, 2	10
3	41.88	Excellent	6, 2	10
4	51.93	Excellent	6, 2	10
5	41.95	Excellent	6, 2	10
6	59.86	Excellent	6, 2	10
7	40.34	Excellent	6, 2	10
8	50.96	Excellent	6, 2	10
9	52.94	Excellent	6, 2	10

Key
Excellent = PSNR > 40dB
Good quality = 30dB< PSNR≤ 40dB
Fair Quality = 20dB < PSNR ≤ 30dB
Poor Quality = PSNR ≤ 20dB

Discussion

Performance Evaluation Metric

We employed the Peak signal-to-noise ratio performance metrics to determine the efficiency of our model. However, only 10 generations were used for the purpose of testing of the system. The formula for calculating the model's Peak Signal to Noise Ratio is given in the following equations:

$$PSNR = 10.\log_{10}(\frac{65025}{MSE}) \quad (11)$$

Performance Evaluation of the System

Our hybrid model was evaluated using PSNR. The computed values of these metrics were captured in Table 1. Our model achieved an accuracy and precision. Hence, it has actively demonstrated that it could serve as an efficiently denoising salt and pepper noise in grey images.

5. Conclusion and Future Scope

The proposed hybrid image denoising model, which integrates fuzzy logic with genetic algorithms, represents a notable advancement in the fields of artificial intelligence and image processing. This comprehensive methodology combines rule-based symbolic reasoning from fuzzy logic with the numerical optimization capabilities of genetic algorithms, demonstrating a mutually beneficial association between distinct computational paradigms.

The model's ability to adapt to varying noise levels and image properties enhances its resilience, rendering it potentially influential in practical domains such as medical imaging and remote sensing. The balance between interpretability and optimization achieved in this hybrid system design furthers our nuanced comprehension of such systems.

The customizability and transferability of the model amplify its usefulness, thereby stimulating future research in the domains of hybrid systems and computational intelligence. In general, the development of this hybrid model not only addresses specific challenges in image denoising but also provides significant techniques and insights to the broader field of artificial intelligence.

Conflict of Interest

Authors declare that they do not have any conflict of interest.

Authors' Contributions

Author-1 (Cookey, Ibiere, Boma): Conceptualized and developed the initial idea, overseeing the entire research process from inception to completion. Including conducting the implementation of the algorithms, supervised the experimentation process, analysed the results, and drafted the manuscript.
Author-2 (Bennett E. O): Surveyed existing algorithms and technologies relevant to the research domain. He collaborated with the lead author in designing experiments, interpreting results, and refining the theoretical framework.
Author-3(Anireh V. I. E): Provided his expertise for the development process, enhancing the robustness and efficiency of the implemented solutions.
Author- 4 (Matthias D.): Researched current existing literatures and formulated the research question for the study. He defined the research objectives and formulating the methodology.
Finally, all authors reviewed and edited the manuscript.

References

[1] Chau, R., Jenkins, M. A., Buchanan, D. D., Ait Ouakrim, D., Giles, G. G., Casey, G., . . Newcomb, P. A., Determining the familial risk distribution of colorectal cancer: a data mining approach. Familial cancer, 15, pp.**241-251, 2016.**

[2] Frénay, B., & Kabán, A., A comprehensive introduction to label noise. Paper presented at the ESANN. **2014.**

[3] S. & Deka, P.C., A basic review of fuzzy logic applications in hydrology and water resources. Appl Water Science, 10, **191, 2020.**

[4] Kumar, A. & Sangaiah, A. K. "A Hybrid Approach to Noise Reduction in Multimedia Data Using Fuzzy Logic and Neural Networks". *International Journal of Multimedia Data Engineering and Management,* Vol.**9**, Issue.**2**, pp.**1-17**, **2018.**

[5] Mirjalili, S., & Mirjalili, S. (2019). Genetic algorithm. Evolutionary algorithms and neural networks: Theory and applications, 43-55.

[6] Nallayarasu, N., Natarajan, P. & Kamaraj, N. "A Hybrid Approach for Image Segmentation Using Fuzzy Logic and Genetic Algorithms". *International Journal of Computer Science and Engineering*, Vol.**5**, Issue.**7**, pp.**123-128**, **2017.**

[7] Ougiaroglou, S., & Evangelidis, G. , Dealing with noisy data in the context of k-NN Classification. Paper presented at the Proceedings of the 7th Balkan Conference on Informatics Conference. **2015.**

[8] Puri, A., & Kumar Gupta, M., Improved hybrid bag-boost ensemble with K-means-SMOTE–ENN technique for handling noisy class imbalanced data. The Computer Journal, Vol.**65**, Issue.**1**, pp.**124-138, 2022.**

[9] Salehi, S. M. R. & Shokrieh, M. R. "A Hybrid Approach for Noise Reduction in Magnetic Resonance Images Using Fuzzy Logic and Genetic Algorithms". *Journal of Biomedical Science and Engineering,* Vol.**9**, Issue.**8**, pp.**409-419**, **2016.**

[10] Wang, H., Huang, L., Zhang, C., & Li, Z., Hybrid Deep Learning and Fuzzy Clustering for Image Denoising. IEEE Transactions on Fuzzy Systems, Vol.**28**, Issue.**4**, pp.**673-682, 2019.**

[11] Ye, Z. Pan, L. & Zhang, Z. "A Hybrid Approach for Image Retrieval Using Convolutional Neural Networks and Genetic Algorithms". *Multimedia Tools and Applications*, Vol.**79**, Issue.**37**, pp.**27693-27714**, **2020.**

[12] Yuan, Z., Wang, Z. & Zhang, D. "A Hybrid Approach for Image Denoising Using Deep Learning and Non-Local Means Filter". *IEEE Access*, Vol.**6**, pp.**69756-69768**, **2018.**

[13] Zhao, W., Wang, H. & Zhang, Y. "A Hybrid Approach for Image Retrieval Using Fuzzy Logic and Genetic Algorithms". *Journal of Multimedia*, Vol.**11**, Issue.**7**, pp.**1448-1457**, **2016.**

[14] Zhang, S., Wang, Y. & Liu, Y. "A fuzzy logic-based diagnostic system for lung cancer". *International Journal of Fuzzy Systems*, Vol.**23**, Issue.**4**, pp.**1219-1229**, **2021.**

[15] Zhao, W., Huang, J. & Zhang, W. (2019). "A Hybrid Approach for Cloud Image Storage Using Convolutional Neural Networks and Genetic Algorithms". *Multimedia Tools and Applications,* Vol.**78**, Issue.**23**, pp.**33711-33724**, **2019.**

AUTHORS PROFILE

Cookey, Ibiere Boma pursued her B.Sc. degree in Computer Science at Rivers State University of Science and Technology, Nigeria in 2006. M.Sc. in Computer Science at University of Glamorgan, Southwales, United Kingdom in 2013. PhD (in View) at the Rivers State University, Port-Harcourt, Rivers State, Nigeria 2024. She is a researcher, a member of the Computer Professionals of Nigeria (CPN), a Certified Project Manager (PRINCE2), has a Diploma in Data Analysis and a PhD student at the said university. She has published four (4) research papers in both local and international journals. Her research works focus on Artificial Intelligence, Machine Intelligence Systems, and Datamining.

E. O. Bennett graduated with a B.Sc. degree in Computer Science from Rivers State University, Rivers State, Port Harcourt, Nigeria in 1998, MSc and PhD at the University of Port Harcourt in 2008 and 2014 respectively. Currently he is an Associate Professor & a Lecturer in the Department of Computer Science, Rivers State University, Port-Harcourt. He is a member of the Computer Professionals of Nigeria (CPN). He has published over 50 research papers in reputed international journals. His research works focuses on Algorithms, Parallel, Distributed & Intelligent Computing.

V.I.E Anireh pursued B.Sc. degree in Computer Science at the University of Nigeria. MSc, & PhD at the University of at the University of Port Harcourt. He is currently an Associate Professor and a Lecturer in the Department of Computer Science, Rivers State University, Port-Harcourt. He is a researcher fellow and a member of IEEE. He has many scholarly publications in both local and international journals. His main research work focuses on Artificial Neural Networks, Machine Learning, Computer Networks, IoT, & Big Data.

D. Matthias pursued B.Sc. in Mathematics/Computer Science at the University of Port Harcourt, Rivers State, M.Tech. from the Federal University of Technology Owerri, Nigeria and PhD from The Rivers State University of Science and Technology, Port Harcourt, Nigeria. He is currently an Associate Professor in the Department of Computer Science, Rivers State University formally Rivers State University of Science and Technology. He is a member of Computer Professional of Nigeria (CPN) since 2010 and a life member of Nigeria Computer Society (NCS) since 2012. He has many scholarly publications in reputed local and international journals. His main research work focuses on Scientific Computing and Database System, Software Development.

