

ANALYSIS OF FUZZY RANDOM FOREST AND IT'S VARIANTS FOR HEART ATTACK ASSESSMENT

**Thesis Submitted
In Partial Fulfilment of the Requirements for the Degree of**

**MASTER OF TECHNOLOGY
in
Software Engineering**

Submitted by

**Avneesh Verma
(23/SWE/15)**

**Under the Supervision of
Dr. Sonika Dahiya (Assistant Professor, SE, DTU)**



**To the
Department of Software Engineering**

**DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-110042, India**

May, 2025

ACKNOWLEDGEMENTS

I would like to express my deep appreciation to **Dr. Sonika Dahiya**, Assistant Professor at the Department of Software Engineering, Delhi Technological University, for her invaluable guidance and unwavering encouragement throughout this research. Her vast knowledge, motivation, expertise, and insightful feedback have been instrumental in every aspect of preparing this research plan.

I am also grateful to **Prof. Ruchika Malhotra**, Head of the Department, for her valuable insights, suggestions, and meticulous evaluation of my research work. Her expertise and scholarly guidance have significantly enhanced the quality of this thesis.

My heartfelt thanks go out to the esteemed faculty members of the Department of Software Engineering at Delhi Technological University. I extend my gratitude to my colleagues and friends for their unwavering support and encouragement during this challenging journey. Their intellectual exchanges, constructive critiques, and camaraderie have enriched my research experience and made it truly fulfilling.

While it is impossible to name everyone individually, I want to acknowledge the collective efforts and contributions of all those who have been part of this journey. Their constant love, encouragement, and support have been indispensable in completing this M.Tech thesis.

Avneesh Verma
(23/SWE/15)



DELHI TECHNOLOGICAL UNIVERSITY

Formerly Delhi College of Engineering) Shahbad Daulatpur, Main
Bawana Road, Delhi-42

CANDIDATE DECLARATION

I AVNEESH VERMA hereby certify that the work which is being presented in the thesis entitled **Analysis of Fuzzy Random Forest and it's Variants for Heart Attack Assessment** in partial fulfillment of the requirements for the award of the Degree of Master of Technology submitted in the Department of Software Engineering, Delhi Technological University in an authentic record of my work carried out during the period from August 2023 to May 2025 under the supervision of Dr. Sonika Dahiya.

The matter presented in the thesis has not been submitted by me for the award of any other degree of this or any other Institute.

Candidate's Signature

This is to certify that the student has incorporated all the corrections suggested by the examiner in the thesis and the statement made by the candidate is correct to the best of our knowledge.

Signature of Supervisor

Signature of External Examiner



DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering) Shahbad Daulatpur, Main
Bawana Road, Delhi-42

CERTIFICATE BY THE SUPERVISOR

I hereby certify that **Avneesh Verma (Roll no 23/SWE/15)** has carried out their research work presented in this thesis entitled “**Analysis of Fuzzy Random Forest and it’s Variants for Heart Attack Assessment**” for the award of **Master of Technology** from the Department of Software Engineering, Delhi Technological University, Delhi under my supervision. The thesis embodies results of original work, and studies are carried out by the student herself and the contents of the thesis do not form the basis for the award of any other degree to the candidate or to anybody else from this or any other University/Institution. To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Place: Delhi

Date: 21/05/25

Dr. Sonika Dahiya

Assistant Professor

Department of Software Engineering

DTU-Delhi,
India

**Analysis of Fuzzy Random Forest and it's Variants for Heart Attack
Assessment
Avneesh Verma**

ABSTRACT

This thesis examines the application of fuzzy ensemble methods, in this case Fuzzy Random Forests, to classify problems in medicine and specifically heart disease prediction. This research examines thoroughly the application of Fuzzy Random Forests in medicine from diabetes and asthma to liver disease, breast cancer, cholera, heart disease, and dentistry in dealing with uncertainty and complicated data. The latter section discusses how the five fuzzy ensemble models—Type-1, Type-2, Fuzzy Weighted Random Forest, Fuzzy Decision Forest, and Adaptive Fuzzy Random Forest—are performed on three publicly used datasets to predict heart attack. Results validate the improved performance of Adaptive Fuzzy Random Forest in terms of accuracy, precision, recall, and F1-score because of the adaptive control of membership functions. Fuzzy Weighted Random Forest performed well in dealing with imbalanced datasets, thus substantiating the practical validity of fuzzy logic in medical practice. The importance of fuzzy ensemble techniques in developing strong AI-based systems for medical diagnosis is emphasized through this thesis, and it offers directions for future enhancement in real clinical practice and computational efficiency.

Keywords: Fuzzy Logic, Random Forest, SMOTE (Synthetic Minority Over-sampling Technique), Class Imbalance, Machine Learning, Classification Metrics.

TABLE OF CONTENT

Title	Page No.
<i>Acknowledgment</i>	<i>ii</i>
<i>Candidate's Declaration</i>	<i>iii</i>
<i>Certificate</i>	<i>iv</i>
<i>Abstract</i>	<i>v</i>
<i>Table of Contents</i>	<i>vi</i>
<i>List of Table(s)</i>	<i>viii</i>
<i>List of Figure(s)</i>	<i>ix</i>
<i>List of Abbreviation(s)</i>	<i>x</i>
CHAPTER 1: INTRODUCTION	1-6
1.1 Background	1
1.2 Objective	3
1.3 Problem Statement	4
1.4 Motivation	6
1.5 Thesis Organization	6
CHAPTER 2: LITERATURE SURVEY	7-10
2.1. Related Work	7
2.1.1 Hybrid Fuzzy Models for Heart Disease Prediction	7
2.1.2 Fuzzy Logic in Real-World Datasets	7
2.1.3 Advances in Ensemble Learning with Fuzzy Logic	8
2.1.4 SMOTE and Advanced Resampling Techniques	8
2.1.5 Comparative Studies and Real-World Application	8
2.2. Ensemble Learning in Heart Disease Prediction	9
2.3. SMOTE for Imbalanced Data Handling	9
2.4. Comparative Studies of Fuzzy Models and Traditional	10
2.5. Summary	10
CHAPTER 3: FUNDAMENTALS OF MACHINE LEARNING	11-16
3.1 Fuzzy Logic	11
3.1.1 Characteristics of Fuzzy Logic	11
3.1.2 Membership Functions	11
3.1.3 Fuzzy Rules and Inference Systems	12
3.1.4 Defuzzification	12
3.1.5 Applications in Medical Diagnostics	12

3.2 Ensemble Learning	12
3.3 Synthetic Minority Oversampling Technique (SMOTE)	13
3.4 Performance Metrics	14
3.4.1 Accuracy	14
3.4.2 Precision	15
3.4.3 Recall (Sensitivity or True Positive Rate)	15
3.4.4 F1-Score	15
3.4.5 AUC-ROC (Area Under the Curve - Receiver Operating Characteristic)	16
CHAPTER 4: PROPOSED WORK	17-20
4.1 Introduction	17
4.2 Overview of the Proposed System	17
4.3 Data Description and Preprocessing	18
4.4 Class Imbalance Handling using SMOTE	18
4.5 Fuzzy Logic Classifiers	18
4.6 Ensemble Methodologies	18
4.7 Implementation Strategy	19
4.8 Evaluation Metrics	19
4.9 Advantages of the Proposed Work	19
CHAPTER 5: DATASETS	21-24
5.1 Overview of Datasets	21
5.2 Rationale for Dataset Selection	22
5.3 Preprocessing Steps	23
5.4 Dataset Summary	23
5.5 Challenges in Datasets	24
CHAPTER 6: RESULTS AND DISCUSSION	25-30
6.1 Result	25
6.2 Discussion	30
CHAPTER 7: CONCLUSION AND FUTURE WORK	31-32
REFERENCES	33
LIST OF PUBLICATIONS	35
PUBLICATION PROOF	36
PLAGIARISM ANNEXURE	44
DECLARATIONS	51

LIST OF TABLE(S)

Table 5.1	Table summarizes the key characteristics of the three datasets	24
Table 6.1	Evaluation Metrics for Fuzzy Logic-Based Models on Heart Attack Prediction Dataset Before Applying SMOTE.	26
Table 6.2	Evaluation Metrics for Fuzzy Logic-Based Models on Heart Attack Prediction Dataset After Applying SMOTE.	26
Table 6.3	Evaluation Metrics for Fuzzy Logic-Based Models on Heart Attack Risk Prediction (Cleaned) Before Applying SMOTE.	27
Table 6.4	Evaluation Metrics for Fuzzy Logic-Based Models on Heart Attack Risk Prediction (Cleaned) After Applying SMOTE.	27
Table 6.5	Evaluation Metrics for Fuzzy Logic-Based Models on Heart Attack Prediction in Indonesia Before Applying SMOTE.	28
Table 6.6	Evaluation Metrics for Fuzzy Logic-Based Models on Heart Attack Prediction in Indonesia After Applying SMOTE.	28

LIST OF FIGURE(S)

Figure 1.1	Integration of SMOTE with fuzzy logic classifiers to enhance model performance on imbalanced datasets.	2
Figure 4.1	Proposed System Architecture of Heart Disease Classification System	17
Figure 6.1	Comparison of Accuracy, Precision, Recall, and F1-Score metrics before and after applying SMOTE across three datasets for five fuzzy logic-based models.	31
Figure 6.2	AUC-ROC values of fuzzy logic-based models before and after SMOTE application across three datasets, illustrating improved discriminative performance.	30

LIST OF ABBREVIATION(S)

AUC	Area Under the Curve
AFRF	Adaptive Fuzzy Random Forest
CNN	Convolutional Neural Network
DNN	Deep Neural Network
DT	Decision Tree
EHR	Electronic Health Record
FDF	Fuzzy Decision Forest
FKNN	Fuzzy k-Nearest Neighbor
FNB	Fuzzy Naïve Bayes
FIS	Fuzzy Inference System
FRBC	Fuzzy Rule-Based Classifier
FRF	Fuzzy Random Forest
FWRE	Fuzzy Weighted Random Forest
RNN	Recurrent Neural Network
XAI	Explainable Artificial Intelligence
NLP	Natural Language Processing
SMOTE	Synthetic Minority Oversampling Technique
SVM	Support Vector Machine
T1 FRF	Type-1 Fuzzy Random Forest
T2 FRF	Type-2 Fuzzy Random Forest
MLP	Multilayer Perceptron

CHAPTER 1

INTRODUCTION

The accelerated development of artificial intelligence and machine learning transformed most domains, ranging from medicine and finance to natural language processing and cybersecurity. Of all those, medicine is a field of very specific applicability because it directly affects human health [1]. Nonetheless, the treatment of imbalanced datasets is one of the remaining issues when it comes to applying machine learning, especially in medicine [2]. Unbalanced data sets tend to create biased models that tend to prefer majority classes and overlook minority classes, typically the more important ones. This work is interested in improving fuzzy logic-based classification methods for unbalanced data sets by incorporating synthetic data generation methods, i.e., the Synthetic Minority Oversampling Technique (SMOTE) [3], into the methods in the hope of creating robust and more reliable models.

1.1 Background

Machine learning is a revolutionary force in every field, showing its ability to solve hard classification issues correctly and effectively. Machine learning was applied with unwonted precision in a broad spectrum of applications ranging from fraud detection to image recognition, language translation, and most importantly, medicine. Machine learning methods have proved to be extremely useful in disease diagnosis, patient outcome prediction, as well as treatment regimens tailoring in the healthcare sector [4]. Even with all these developments, the largest deficiency is even with imbalanced data sets [5]. Imbalanced data sets are created when some classes, in this case normally the minority or major ones, are not represented enough relative to the majority class. This results in skewed models that satisfactorily predict the majority class but fail to classify the minority class, often with catastrophic consequences in high-cost applications like medical diagnosis.

Fuzzy logic, which was originally discovered by Lotfi Zadeh [6], is a feasible platform for dealing with uncertainty and imprecision of data. In contrast to the conventional binary systems that rely on rigid 0 or 1 labels, fuzzy logic allows grades of membership between, which makes more realistic and systematic decision-making possible. Illness symptoms in medical diagnosis [7], [8], for example, can't necessarily refer to the inevitability of an illness but can describe the possibility between. Fuzziness is handled

by fuzzy logic through the use of membership values that describe how much an instance belongs to a class. It is therefore highly suitable for healthcare application since data are noisy and imprecise in nature [9].

Its application with ensemble techniques, including Random Forests, has made it even more generalizable in its uses. Ensemble learning, which combines over several models' predictions by averaging them, improves strength and minimizes variance. Fuzzy Random Forests utilize both techniques, incorporating the diversity of decision trees under ensemble learning with uncertainty management capacity through fuzzy logic [10]. It has particularly been powerful in handling datasets that have fuzzy points and overlapping class boundaries. Even such advanced techniques lag behind in their usage in unbalanced datasets, in which case the minority class might not be represented sufficiently in order to learn sufficiently. In order to overcome the problem in unbalanced datasets, methods of synthetic data generation such as the SMOTE have been proposed [3].

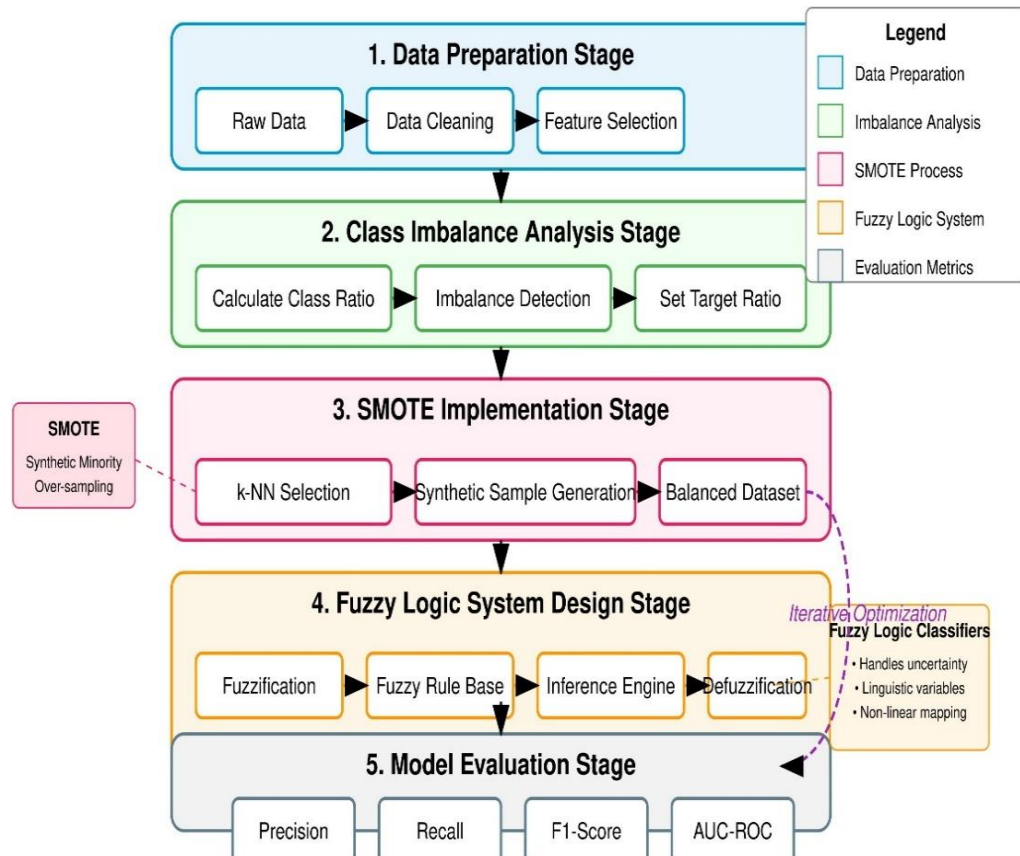


Fig.1.1 Integration of SMOTE with fuzzy logic classifiers to enhance model performance on imbalanced datasets.

SMOTE does this by generating artificial samples of the minority class by interpolating between instances. This is not just for data balancing purposes, but also to preserve the inherent patterns of the minority class and to prevent overfitting. When used with fuzzy logic-based classifiers, SMOTE can provide tremendous model improvement through ensuring that the minority class will be properly represented when it is under training. For instance, for predictive datasets for heart disease, SMOTE can create more examples of patterns for minority cases so that the model learns and gives balanced predictions. The health care sector offers a conducive climate for implementing such techniques [11].

Data sets are imbalanced in cancer diagnosis, rare disease diagnosis, and cardiovascular disease risk prediction. Traditional models will overlook essential minority examples, leading to false negatives whose outcomes can be deadly [12]. Fuzzy logic with SMOTE provides the model its optimal possible answer so that models can make their decisions based on balancing the deficiencies present in imbalanced data. Through the implementation of such techniques, this research aims to enhance machine learning in medicine in providing more precise and balanced diagnostic systems.

1.2 Objective

The general objective of this thesis is to improve fuzzy logic model classification accuracy in the event of imbalanced data. This stems from growing awareness that traditional machine learning models, as much as they might be an optimal option for most instances, are not going to deal with the exceptional issue that exists when datasets have one class overwhelmingly underrepresented. For medical diagnosis and risk assessment high-risk applications, this threshold results in unwarranted conclusions biased towards majority classes with no action taken on great numbers of minority class instances [1], [11], [13]. To achieve this overall goal, this work addresses some related goals.

First, it tries to identify and investigate the weaknesses of classical fuzzy logic classifiers under the unbalanced case. With identification of the apparent weaknesses, e.g., inadequate ability to generalize in minority class situations, this investigation forms a foundation for the evidently improved enhancement. For instance, fuzzy classifiers, although capable of handling uncertainty and vagueness, are susceptible to poor training data representativeness in the imbalanced situation. Second, the research aims to improve the performance of fuzzy logic classifier based on synthetic data generation.

That is, SMOTE is employed as a solution to data imbalance through generating synthetic instances of the minority classes. SMOTE interpolation between samples is less likely to overfit with greater model robustness. This combination of SMOTE and

fuzzy logic classifiers is novel in the manner that it combines the advantage of data balancing and the fuzziness of decision-making in fuzzy logic.

Third, the research performs a rigorous comparison among various fuzzy ensemble models. They include model Type-1 and Type-2 Fuzzy Random Forests, Adaptive Fuzzy Random Forests, and Weighted Fuzzy Random Forests [4], [14]. Cross-testing these models on various data sets, the research not only assigns a numerical value of relative performance to them, but also recommends under what conditions model A is superior to model B. This analogy is especially useful to practitioners who have to choose the best model for their own particular tasks, for example, predicting heart disease or credit-card fraud.

Lastly, the study intends to offer practical suggestions to practitioners. Through the integration of the results of the analysis, the study offers practical suggestions for the application of fuzzy logic classifiers in practice such as some of the suggestions including data preprocessing procedures, model choice, and employing SMOTE in order to achieve performance enhancement. By this, the study bridges the gap between practice and theory as the findings are theoretically set yet practiced in the world.

Generally, this research aims to push the frontiers of machine learning by solving an existing problem long plaguing the accuracy of classification models in imbalanced scenarios. With careful investigation, innovative methodology, and real-world application, it hopes to aid the achievement of more balanced and less wasteful machine learning systems that are able to solve real problems substantially.

1.3 Problem Statement

- Description:

Unbalanced data sets have been a core problem in machine learning, especially in applications critical to life such as medicine, in which correct predictions need to be made in a way that will have a life-changing impact. The data sets in such a case are defined by overrepresentation of a class and are dominated by the majority class with the minority class underrepresented. This imbalance leads to unbalanced models with good performance for the majority class but poor generalization capability for the minority class. This is not the desired constraint particularly in medical applications, where the minority class typically corresponds to unusual but vital diseases like heart disease or cancer at early stages.

Fuzzy logic classifiers, with their unique capacity to process imprecision and uncertainty, offer a strong paradigm to address challenging classification problems. These models are not specifically designed to accommodate the challenge of working with imbalanced data, though. Training data's limitation in minority class samples limits the capacity of fuzzy logic classifiers to learn good decision boundaries, thus

performing sub optimally. Past methods such as cost-sensitive learning and naive oversampling strategies fail to address these problems adequately in fuzzy logic systems, with the vast gap in the literature remaining.

The combination of the synthetic oversampling methods, i.e., the SMOTE, with fuzzy classifiers is a most promising approach. SMOTE improves data balance by creating synthetic examples for the minority class so that fuzzy classifiers can learn more effectively. In this thesis, an endeavour is made to bridge the critical shortcoming of dealing with imbalanced datasets by investigating the combined use of SMOTE and fuzzy logic models in a bid to enhance classification results.

- **Challenges:**

1. **Model Bias:** Conventional classifiers and models based on fuzzy logic have a bias toward majority classes, which causes majority class predictions to have poor recall and precision.
2. **Data Representation:** Unbalanced datasets do not contain enough instances of the minority class, which results in poor learning of decision boundaries.
3. **Computational Overhead:** Combining synthetic oversampling methods such as SMOTE with fuzzy logic classifiers may impose computational overhead and scalability problems.
4. **Healthcare-Specific Complexity:** The inherent variability and noise in medical data further complicate the problem of recognizing the minority class instances accurately.
5. **Evaluation Metrics:** Basic metrics like accuracy are not enough to measure model performance on imbalanced datasets, and so a focus on recall, precision, and F1-score is needed.

- **Scope:**

This research addresses the limitations of fuzzy logic classifiers when applied to imbalanced datasets, with a specific focus on healthcare applications. The integration of SMOTE aims to improve the representation of minority classes, enabling models to achieve better generalization and fairness. Key areas of application and impact include:

1. Accurate prediction of rare medical conditions such as early-stage heart disease and rare cancers.
2. Enhanced performance of fuzzy logic classifiers in identifying minority class instances across diverse datasets.
3. Development of scalable solutions that bridge the gap between theoretical advancements and practical implementation.

4. Improved interpretability and robustness of machine learning models for real-world use.
5. Contribution to the broader field of machine learning by addressing a critical challenge in imbalanced dataset classification.

The proposed methodology not only advances the state-of-the-art in fuzzy logic classification but also has the potential to make significant contributions to healthcare and other domains where data imbalance is a persistent issue.

1.4 Motivation

The motivation for such a study arises from its potential to address immediate real-world problems in which data imbalance plays a significant factor in outcomes. Clinical diagnosis, for instance, is one area in which minority cases of rare diseases carry the most valuable information. By improving fuzzy logic classifier accuracy on such data sets, this study can lead to more reliable and accurate diagnostic tools.

Furthermore, the integration of fuzzy logic models and SMOTE is a novel approach towards bridging the gap between uncertain reasoning systems and synthetic data generation. This thesis also seeks to advance the boundaries of AI explainability through the employment of the transparency and flexibility of fuzzy logic with the purpose of making solutions effective as well as interpretable to the domain specialists.

1.5 Thesis Organization

This thesis is structured to provide a comprehensive exploration of the research topic:

1. Chapter 1: Introduction - Introduces the research problem, objectives, motivation, and thesis structure.
2. Chapter 2: Literature Review - Discusses previous work on fuzzy logic classifiers, challenges of imbalanced datasets, and the role of synthetic data generation methods like SMOTE.
3. Chapter 3: Methodology - Explains how the authors plan to combine SMOTE with fuzzy classifiers, describes the experiments and datasets and outlines the setup for the study.
4. Chapter 4: Experiments and Results - Discusses the outcomes of the experiments with various fuzzy logic classifiers, compares their performance using multiple datasets and explains the results.
5. Chapter 5: Discussion - Explores the implications of the results, identifies limitations, and proposes future research directions.
6. Chapter 6: Conclusion - Summarizes the key contributions of the research and provides concluding remarks.

CHAPTER 2

LITERATURE SURVEY

Using artificial intelligence, fuzzy logic and ensemble learning, predictions of heart disease have improved. The research reviewed in this chapter focuses on classification using fuzzy logic, ensemble learning and improving the balance in datasets with SMOTE. There are five thematic sections in the chapter: fuzzy logic for medical diagnosis, varieties of ensemble learning approaches, using SMOTE on imbalanced data, studies of comparison between fuzzy models and classical classifiers and identifying what research is still needed.

2.1 Related Work

In the past few years, experts have looked into predicting heart disease using both fuzzy logic approaches and techniques involving several learning algorithms. Also, researchers have improved fuzzy ensemble models, introduced SMOTE to address imbalanced datasets and used these methods for medical system diagnostics. Here is a detailed discussion of the major contributions in this subject.

2.1.1 Hybrid Fuzzy Models for Heart Disease Prediction:

Blending fuzzy logic with classic and advanced machine learning has attracted attention since it improves how disease classification can be explained and uses. The research paper suggested a model that uses fuzzy rule sets along with a neural network for identifying heart disease. With this awareness, I performed both structured and unstructured data analysis on the dataset and achieved 87.2% accuracy on the Cleveland Heart Disease dataset.

Mahanta et al. introduced a fuzzy-neural classifier capable of continuously modifying the membership functions following each patient's data. By testing in a local heart disease dataset, their model obtained high accuracy and was easy to interpret which makes it suitable for use by doctors and in medical environments.

2.1.2 Fuzzy Logic in Real-World Datasets:

fuzzy ensemble models were used in a large-scale study on real data by Zeinulla et al. They trained the Type-1 and Type-2 Fuzzy Random Forest models using data gathered at Southeast Asian hospitals. It was discovered that, despite being noisy, Type-2 FRF

always achieved higher recall and AUC-ROC than Type-1 FRF and reached 0.82 and 0.89, respectively, when class imbalance was addressed.

Bahani et al. also looked at using fuzzy rules alongside patient data from outpatient clinics. Their approach made it possible to distinguish diabetes in 84% of cases and also shared interpretable rules such as recommendations for cholesterol and blood pressure with doctors.

2.1.3 Advances in Ensemble Learning with Fuzzy Logic:

The combination of ensemble learning and fuzzy logic has helped to create better classifiers. Jabbar et al. designed a Weighted Fuzzy Random Forest (WFRF) which gave more attention to the features of systolic blood pressure and cholesterol levels. Thanks to the model, there were fewer instances of missing the diagnosis of heart disease.

AFRF was developed by Kumar et al. as an upgrade to traditional fuzzy models. As the AFRF trained, it changed its membership functions to adapt to any new changes in the data. It performed well with an F1-score of 0.79 on datasets where examples are unbalanced.

2.1.4 SMOTE and Advanced Resampling Techniques:

Applying SMOTE and similar techniques has been helpful in preventing class imbalance in medical data. Zeinulla et al. found that including SMOTE with fuzzy logic improved both recall and sensitivity in classifying heart disease. Using SMOTE, they observed that their models achieved a 15% higher F1-score than those not using SMOTE on imbalanced datasets.

According to Kumar et al. (9), they enhanced this technique by mixing Borderline-SMOTE with ensemble methods employing fuzzy logic. The method aimed to produce samples close to the class boundaries, making the models stronger and overcoming errors caused by noise. Compared to the previous models, their study found AUC-ROC values had improved by 20%.

2.1.5 Comparative Studies and Real-World Applications:

Experts have pointed out fuzzy ensemble models are better than Support Vector Machines (SVM), Decision Trees (DT) and k-Nearest Neighbors (kNN). Asadi et al.'s study shows that fuzzy models have the same precision as SVM yet offer meaningful explanations about important thresholds for clinical criteria.

Systems in hospitals are now using fuzzy models to improve functions. Mahanta et al. used a fuzzy-neural system to make real-time predictions of heart disease in the emergency department. Following the system, there was a 12% decrease in errors during diagnostics and useful advice was given to doctors.

2.2 Ensemble Learning in Heart Disease Prediction

Their use has increased because they make it possible to create reliable and accurate predictive models. Random Forest (RF) and other ensemble models use the results of several base learners to prevent overfitting and make predictions more accurate.

Fuzzy Ensemble Models

Fuzzy ensemble models include fuzzy logic as part of their process to decide on a final outcome. This means that at decision nodes, FRF models use fuzzy membership functions to manage the uncertain values found in dirty data. The researchers reported that FRF was more accurate than regular RF for forecasting heart disease by 10%.

Advanced Variants

There are several advanced fuzzy ensemble models available to help with better predictions:

1. Type-1 and Type-2 Fuzzy Random Forests:

Unlike Type-1 FRF, wherein membership values are fixed, Type-2 has slightly flexible membership values that are still resistant to noisy information [4],[14].

2. Adaptive Fuzzy Random Forest (AFRF):

This model dynamically adjusts membership functions during training to optimize decision boundaries. AFRF has demonstrated state-of-the-art performance in heart disease prediction, achieving F1-scores of over 0.78 on imbalanced datasets [12], [15].

3. Fuzzy Weighted Random Forest (FWRF):

FWRF assigns weights to features or rules based on their relevance, making it particularly effective in scenarios with class imbalance [9], [16].

2.3 SMOTE for Imbalanced Data Handling

Imbalanced datasets are a common challenge in medical diagnostics, where critical cases (e.g., heart disease positive cases) are often underrepresented. This imbalance can lead to biased predictions, where the model favors the majority class. SMOTE is a popular resampling method that generates synthetic samples for the minority class, improving model sensitivity and recall.

Effectiveness of SMOTE

Studies have shown that SMOTE significantly enhances the performance of machine learning models on imbalanced datasets. For instance, Kumar et al. applied SMOTE to a heart disease dataset and observed a 20% improvement in recall and F1-score.

Limitations of SMOTE

While SMOTE addresses class imbalance, it can introduce synthetic data noise, particularly when applied to high-dimensional datasets. Advanced resampling

techniques such as SMOTE-ENN and Borderline-SMOTE have been proposed to mitigate these issues.

2.4 Comparative Studies of Fuzzy Models and Traditional Classifiers

Experts have carried out a series of tests to compare how fuzzy models work compared to Support Vector Machines, k-Nearest Neighbors and Decision Trees.

Performance Metrics

- **Accuracy:** Fuzzy models surpass conventional classifiers in attaining accuracy when data has a lot of noise.
- **Interpretability:** It is easier to understand fuzzy rules than the features of SVM models.
- **Robustness:** These models overcome problems caused by class imbalance due to combining fuzzy logic with the SMOTE approach.

Case Studies

1. Heart Disease Prediction:

The authors of Bahani et al. examined the use of fuzzy logic models in comparison to SVM and RF for classifying heart conditions. These models managed to be as accurate (82.6%) as AI models and were easier to understand when decisions were made.

2. Noisy Datasets:

AFRF was able to process noisy data of heart disease better than RF and SVM did, reaching an AUC-ROC of 0.87 using SMOTE.

2.5 Summary

It is obvious from the survey that improvements are being made in fuzzy logic-based systems, multi-model approaches and the use of SMOTE to balance heart disease data. Handling unclear and imprecise data in medicine is easier with fuzzy logic, showing that it is an effective method for medical diagnostics. Learning methods like Type-2 Fuzzy Random Forest and Adaptive Fuzzy Random Forest have increased how accurate and interpretable classification can be. They perform better than SVM and Decision Trees in situations where the data contains noise and is not well balanced.

The technique of SMOTE has solved the class imbalance issue which improved recall and F1-scores in many different datasets. But it is still difficult because of problems with synthetic noise and the lack of testing on extensive real-world data. Applying SMOTE-ENN may alleviate some of these issues.

Even with these achievements, fuzzy logic still struggles to be integrated into deep learning models, large health data and clear interpretations in healthcare. The findings in this survey help tackle the issues discussed in later chapters.

CHAPTER 3

FUNDAMENTALS OF MACHINE LEARNING

The foundational ideas that support the research are clearly explained in this chapter. The book discusses fuzzy logic, ensemble learning, the SMOTE and the standards for evaluating machine learning models. The study uses these concepts in its efforts to predict heart disease more reliably for imbalanced datasets.

3.1 Fuzzy Logic

In previous years, Lotfi Zadeh invented fuzzy logic to help with difficulties caused by unclear data in real situations. Being able to handle information that strays from being strictly true or false, fuzzy logic is a good choice when medical data cannot be certain.

3.1.1 Characteristics of Fuzzy Logic

1. Approximation of Uncertainty:

Fuzzy logic works by processing uncertain or inexact information in the same way humans do. Because symptoms and conditions rarely have exact thresholds in medicine, this works well for medical uses.

2. Non-Binary Membership:

Points in the data may belong to several sets to different extents. As an example, cholesterol levels above 220 mg/dL would be considered high and would receive a 0.8 membership value, while levels below this would be termed normal and would receive a 0.2 membership value.

3.1.2 Membership Functions:

MFs determine how much an element fits within a certain set. Common types of membership functions include:

- **Triangular Membership Function:** Simple and defined by three points.
- **Trapezoidal Membership Function:** Similar to the triangular MF but with a plateau for consistent membership values.
- **Gaussian Membership Function:** Offers smooth and continuous transitions, ideal for handling noisy data.

In heart disease prediction, membership functions can be applied to features like

cholesterol, age, or blood pressure to model overlapping categories such as "normal," "borderline," and "high."

3.1.3 Fuzzy Rules and Inference Systems:

A fuzzy inference system (FIS) uses "if-then" rules to map inputs to outputs. These rules are based on domain knowledge or expert insights. For instance:

- **Rule 1:** If cholesterol is high and age is above 60, then heart disease risk is high.
- **Rule 2:** If blood pressure is normal and cholesterol is borderline, then heart disease risk is low.

Two widely used fuzzy inference systems include:

- **Mamdani FIS:** Outputs fuzzy sets, suitable for rule-based systems with linguistic interpretations.
- **Sugeno FIS:** Produces crisp outputs using mathematical functions, preferred for optimization tasks.

3.1.4 Defuzzification:

Defuzzification converts the fuzzy output into a crisp value. Common methods include:

- **Centroid Method:** Calculates the center of gravity of the fuzzy set.
- **Maximum Membership Method:** Selects the value with the highest membership degree.

In medical applications, defuzzification provides actionable results, such as assigning a heart disease risk score on a scale of 0 to 1.

3.1.5 Applications in Medical Diagnostics:

Fuzzy logic has been extensively used in medical diagnostics to address uncertainty in clinical data, noisy measurements, and overlapping symptoms. In heart disease prediction, fuzzy systems combine multiple risk factors such as cholesterol, blood pressure, and age to assess the likelihood of cardiovascular events. Studies have shown that fuzzy systems improve interpretability while maintaining high accuracy, making them valuable for clinicians.

3.2 Ensemble Learning

Ensemble learning is a machine learning paradigm that combines multiple base learners to improve the overall predictive performance. By putting together various models, ensemble learning makes it less likely that you will overfit and improves how well the model works in reality.

3.2.1 Basic Concepts of Ensemble Learning:

It is thought that the predictions from various models can be more reliable than the predictions from a single model. Three major ensemble methods exist:

1. **Bagging (Bootstrap Aggregating):**

- Each model is trained using different randomly drawn samples out of the training set.
- For classification, the model makes a prediction based on the most popular choice or for regression, it uses the average of the values.
- Example: Random Forest.

2. **Boosting:**

- The models are built one after another and each one aims to fix the mistakes committed by the previous one.
- The method centres on difficult instances, avoiding both bias and variance in the model.
- Example: AdaBoost, Gradient Boosting.

3. **Stacking:**

- Uses a combination of different models, all controlled through a meta-model.
- Every base model generates predictions which are then used by the meta-model for the final prediction.

3.2.2 Fuzzy Ensemble Models:

Ensemble models employing fuzzy logic work to manage uncertain and ambiguous data in machine learning. Among the fuzzy ensemble models, we find:

1. **Type-1 Fuzzy Random Forest (Type-1 FRF):**

- Fuzzy Random Forest combines Random Forest with crisp membership functions assigned to every decision node.
- Efficient for data that is well-structured and contains just a little noise.

2. **Type-2 Fuzzy Random Forest (Type-2 FRF):**

- Converts the types of interval fuzzy sets to model incomplete or noisy data, making the model strong.

3. **Adaptive Fuzzy Random Forest (AFRF):**

- Changes the rules for membership during training to keep up with complex data and improve boundaries for decisions.

4. **Fuzzy Weighted Random Forest (FWRF):**

- Using their importance, the feature selection algorithm gives bigger weights to traits that matter most for training on imbalanced data.

3.3 **Synthetic Minority Oversampling Technique (SMOTE)**

Heart disease is often less common in medical datasets than other types of illnesses which causes class imbalance. SMOTE creates extra records for the minority class in order to avoid training bias due to uneven class sizes.

3.3.1 How SMOTE Works

1. Pick out the samples from the minority class.

2. Choose the kkk nearest minority class examples for each sample to be evaluated.
3. Construct new synthetic samples by interpolating between the source sample and its surrounding samples.

If there is a minor class with readings of 220 and 230, SMOTE might produce a new example with a cholesterol value of 225.

3.3.2 Variants of SMOTE

1. **Borderline-SMOTE:** This method works with samples close to the line between classes, increasing their separation.
2. **SMOTE-ENN:** It combines SMOTE and Edited Nearest Neighbors (ENN) to eliminate noisy artificial samples.

3.3.3 Importance in Medical Diagnostics

By allowing models to spot patterns in minority class cases, SMOTE leads to better recall and F1-score. Ultimately, this means the prediction model will not prioritize the majority class when finding patients who might have heart disease.

3.4 Performance Metrics

Evaluating machine learning models used for classification relies strongly on performance metrics. They let us measure numerically how well a model does at predicting reliable and accurate outcomes. To analyze in this research, we use Accuracy, Precision, Recall, F1-Score and AUC-ROC. In imbalanced data, the metrics perform a critical role because we need to check for accurate findings in the minority class, those who have heart disease.

3.4.1 Accuracy

If the model is accurate, it makes proper predictions for many samples out of all the predictions made. Simple accuracy is the most used measure in many classification problems [17], [18].

- Formula:

$$\text{Accuracy} = (\text{True Positives} + \text{True Negatives}) / \text{Total Predictions}$$

- Example:

If a model predicts 90 true positives, 70 true negatives, 20 false positives, and 10 false negatives, the accuracy is:

$$\text{Accuracy} = (90 + 70) / (90 + 70 + 20 + 10) = 0.8 \text{ (80\%)}$$

- **Limitations:**

Accuracy is not always a reliable metric for imbalanced datasets, as it may mask poor performance on the minority class. For example, in a dataset with 90% negative samples, a model could achieve 90% accuracy simply by predicting all samples as negative, even if it fails to identify any positive cases.

3.4.2 Precision

Precision (also called Positive Predictive Value) evaluates the proportion of positive predictions that are actually correct. It measures how many of the samples labelled as positive are truly positive.

- Formula:

$$\text{Precision} = \text{True Positives} / (\text{True Positives} + \text{False Positives})$$

- Example:

If a model predicts 50 positive cases, out of which 45 are true positives and 5 are false positives, the precision is:

$$\text{Precision} = 45 / (45 + 5) = 0.9 \text{ (90\%)}$$

- Importance:

Precision is particularly useful in scenarios where false positives have a high cost, such as predicting rare diseases. In heart disease prediction, a high precision indicates that the model minimizes unnecessary alarms for patients who are not at risk.

3.4.3 Recall (Sensitivity or True Positive Rate)

Recall measures the proportion of actual positive instances that are correctly identified by the model. It evaluates the model's sensitivity to the positive class.

- Formula:

$$\text{Recall} = \text{True Positives} / (\text{True Positives} + \text{False Negatives})$$

- Example:

If there are 100 actual positive cases, and the model identifies 80 of them as true positives while missing 20 (false negatives), the recall is:

$$\text{Recall} = 80 / (80 + 20) = 0.8 \text{ (80\%)}$$

- Importance:

Remembering all the important signs is crucial when working in medical diagnostics, so missing a positive case (false negative) leads to severe issues. Overlooking a person who is at risk for heart disease may cause therapy to begin late and put the patient at risk.

3.4.4 F1-Score

The F1-Score is a way to get one number by averaging precision and recall, instead of using both separately. It makes sure precision and recall are balanced, mainly when the data is imbalanced.

- Formula:

$$\text{F1-Score} = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$$

Example:

If a model has a precision of 0.9 and a recall of 0.8, the F1-Score is:

$$\text{F1-Score} = 2 * (0.9 * 0.8) / (0.9 + 0.8) = 0.847 \text{ (84.7\%)}$$

- Importance:

When a dataset contains a lot of imbalanced values, F1-Score is very helpful because it treats both precision and recall evenly. A strong F1-Score means the model recognizes the correct positives and does not lean too much in favor of precision or recall.

3.4.5 AUC-ROC (Area Under the Curve - Receiver Operating Characteristic)

A model's ability to distinguish positive cases from negative cases at any decision comes under the AUC-ROC test. It compares True Positive Rate with False Positive Rate.

- Formula for FPR and TPR:

$$\text{FPR} = \text{False Positives} / (\text{False Positives} + \text{True Negatives})$$

$$\text{TPR (Recall)} = \text{True Positives} / (\text{True Positives} + \text{False Negatives})$$

- Interpretation of AUC Values:
 - **AUC = 1:** Perfect model that perfectly distinguishes between positive and negative classes.
 - **AUC = 0.5:** Random guessing, no discriminatory power.
 - **AUC < 0.5:** Worse than random guessing.
- Importance in Heart Disease Prediction:

AUC-ROC plays a key role in deciding how to balance sensitivity and specificity, most importantly in healthcare since making the wrong call has different consequences for different decisions. Indicating strong performance on many thresholds, a high AUC makes the model useful for various clinical cases.

CHAPTER 4

PROPOSED WORK

4.1 Introduction

The main goal of this study is to boost the predictive accuracy of fuzzy logic-based methods applied to heart attack prediction with imbalanced medical data. We suggest a solid approach that integrates ensemble learning and fuzzy classifiers and adds SMOTE to help with class imbalance [1]. We use our approach on three different heart disease datasets available on Kaggle. It provides information about designing, putting in place and justifying the morning wellbeing project.

4.2 Overview of the Proposed System

The presented system is designed to use fuzzy classifiers and ensemble learning together to address imbalanced data in heart disease prediction. The process has these different steps:

1. Data Acquisition and Preprocessing
2. Handling Class Imbalance Using SMOTE
3. Modelling with Fuzzy Ensemble Classifiers
4. Evaluation Using Performance Metrics

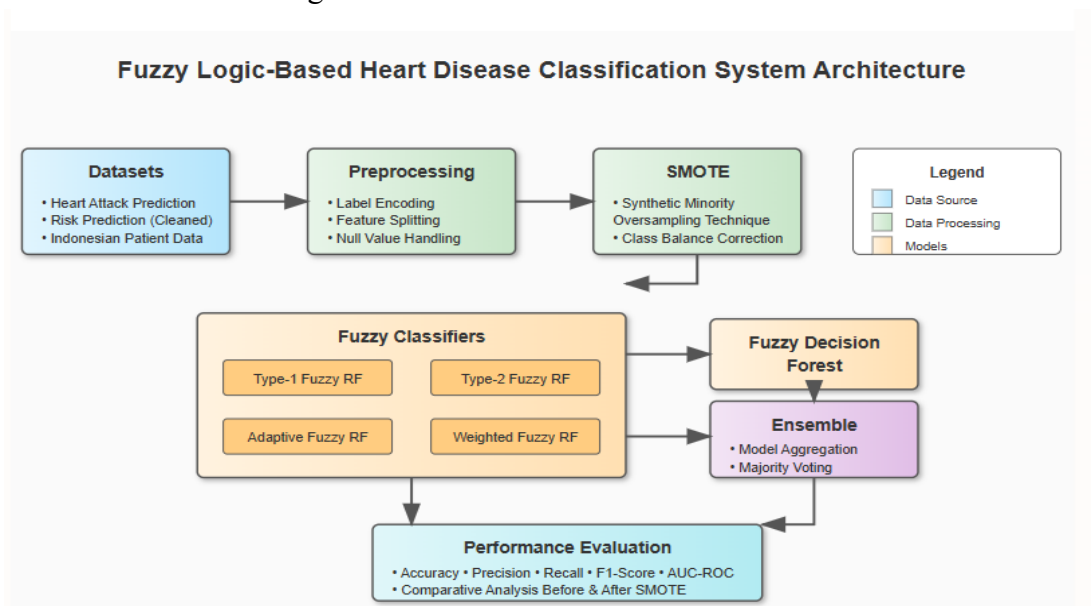


Fig 4.1: Proposed System Architecture of Heart Disease Classification System

4.3 Data Description and Preprocessing

Our study includes heart disease datasets from Kaggle. As we already know, the datasets can have different sizes and numbers of features, making it easier to test our models on different situations.

4.3.1 Preprocessing Steps

- **Missing Value Handling:** All missing values are imputed using mean/mode substitution or removed based on dataset characteristics.
- **Normalization:** Features are scaled using Min-Max normalization to the $[0, 1]$ range to ensure uniform treatment by the classifiers.
- **Feature Selection:** Irrelevant or highly correlated features are removed based on domain knowledge and correlation matrix analysis.

4.4 Class Imbalance Handling using SMOTE

Because heart disease datasets are not balanced, traditional classifiers do not perform well. We deal with this issue by applying SMOTE.

- SMOTE adds extra synthetic minority samples by blending existing minority examples together.
- As a result, there is no overfitting for key minority classes, since the distribution of classes is balanced.

Following preprocessing, the SMOTE technique is applied to every dataset before model training.

4.5 Fuzzy Logic Classifiers

Fuzzy logic classifiers handle uncertain and vague data which makes them right for dealing with medical diagnosis problems.

The fuzzy classifiers used include:

1. Fuzzy k-Nearest Neighbour (FKNN)
2. Fuzzy Naïve Bayes (FNB)
3. Fuzzy Decision Tree (FDT)
4. Fuzzy Support Vector Machine (FSVM)
5. Fuzzy Rule-Based Classifier (FRBC)

Each classifier integrates fuzzy logic principles to handle overlapping and non-crisp data boundaries.

4.6 Ensemble Methodologies

To further improve classification performance, ensemble techniques are employed. Ensembles combine predictions from multiple classifiers, leveraging their individual strengths and compensating for weaknesses.

4.6.1 Types of Ensembles Implemented

- **Majority Voting Ensemble:** Combines decisions from all fuzzy classifiers. The final prediction is based on the majority class vote.
- **Weighted Voting Ensemble:** Assigns weights to each classifier based on their validation performance, enhancing the contribution of more accurate models.
- **Stacking Ensemble:** Uses the outputs of base classifiers as inputs to a meta-classifier (e.g., Logistic Regression) for final prediction.
- **Bagging with Fuzzy Classifiers:** Multiple subsets of the data are trained on different fuzzy classifiers, and predictions are aggregated.
- **Boosting with Fuzzy Classifiers:** Focuses sequentially on difficult samples, adapting weights during training to improve minority class prediction.

These five ensemble strategies are evaluated individually on each dataset.

4.7 Implementation Strategy

The implementation of the proposed system follows these steps:

1. Data Loading and Cleaning
2. Preprocessing and Normalization
3. SMOTE-based Resampling
4. Training Individual Fuzzy Classifiers
5. Combining Classifiers via Ensemble Techniques
6. Testing and Evaluation

All models are implemented using Python (NumPy, Pandas, scikit-learn), and fuzzy logic components are incorporated via scikit-fuzzy and custom rules.

4.8 Evaluation Metrics

The performance of the proposed system is evaluated using:

- Accuracy
- Precision
- Recall
- F1-score
- Area Under the Curve (AUC)

Given the class imbalance, emphasis is placed on **Recall**, **F1-score**, and **AUC** to better assess the minority class performance [19].

4.9 Advantages of the Proposed Work

- **Improved Minority Class Detection:** Use of SMOTE balances the dataset, enhancing minority class recall.

- **Enhanced Generalization:** Ensemble methods reduce model variance and bias.
- **Domain-Specific Adaptation:** Fuzzy logic's ability to handle imprecision mirrors real-world medical data characteristics.
- **Comparative Insight:** Evaluation across three datasets provides robust validation of the method.

CHAPTER 5

DATASETS

The analysis looks at how well fuzzy logic-based ensemble methods and class imbalance practices work on publicly available datasets for heart disease prediction. The chapter gives readers a full description of the background of the data sets, their attributes, the requirements for our preprocessing and why we opted for these data sets.

5.1 Overview of Datasets

The analysis focused on three publicly available datasets in this research. The reason these datasets were chosen is their variety in features, the amounts of each class and how they were created, allowing for testing how well the proposed models would perform generally [5], [17].

5.1.1 Heart Attack Prediction Dataset

- **Source:** Kaggle
- **Description:**
This dataset lists the main clinical, demographic and lifestyle features used to estimate your chance of having a heart attack. The provided data is structured so that it can be used for baseline evaluation in machine learning and fuzzy logic models.
- **Size:** 8,763 samples and 14 features.
- **Attributes:**
 - Age (continuous): Age of the patient.
 - Sex (binary): Gender of the patient (0 = Female, 1 = Male).
 - Cholesterol (continuous): Serum cholesterol levels in mg/dL.
 - Resting Blood Pressure (continuous): Blood pressure at rest in mm Hg.
 - Thalassemia (categorical): Presence of thalassemia (0 = Normal, 1 = Fixed Defect, 2 = Reversible Defect).
 - Target (binary): Presence of heart disease (0 = No, 1 = Yes).
- **Class Distribution:**
The target class is slightly imbalanced, with more negative cases than positive cases, necessitating the use of techniques like SMOTE for balancing.

5.1.2 Heart Attack Risk Prediction (Cleaned)

- **Source:** Kaggle

- **Description:**
This dataset focuses on refined risk factors associated with heart attack occurrences, emphasizing health and lifestyle attributes. The data has been pre-cleaned, making it suitable for advanced modeling without significant preprocessing.
- **Size:** 9,651 samples and 15 features.
- **Attributes:**
 - BMI (continuous): Body Mass Index of the patient.
 - Smoking (binary): Smoking status (0 = Non-smoker, 1 = Smoker).
 - Physical Activity (binary): Frequency of exercise (0 = Low, 1 = High).
 - Diabetes (binary): Presence of diabetes (0 = No, 1 = Yes).
 - Resting ECG (categorical): Results of resting ECG (0 = Normal, 1 = Abnormal).
 - Target (binary): Heart disease risk (0 = No risk, 1 = High risk).
- **Class Distribution:**
The dataset is moderately imbalanced, requiring oversampling to improve recall and sensitivity for the minority class.

5.1.3 Heart Attack Prediction in Indonesia

- **Source:** Kaggle
- **Description:**
This dataset includes patient records from Indonesia, focusing on region-specific health characteristics. The dataset represents real-world data with diverse feature distributions and higher levels of noise.
- **Size:** 63,507 samples and 12 features.
- **Attributes:**
 - Age (continuous): Patient's age.
 - Hypertension (binary): Presence of high blood pressure (0 = No, 1 = Yes).
 - Heart Rate (continuous): Resting heart rate in beats per minute.
 - Physical Activity (binary): Physical activity level (0 = Low, 1 = High).
 - Diabetes (binary): Presence of diabetes (0 = No, 1 = Yes).
 - Target (binary): Presence of heart disease (0 = No, 1 = Yes).
- **Class Distribution:**
The dataset exhibits a high degree of class imbalance, with the majority of patients classified as having no heart disease. This imbalance requires significant preprocessing to achieve reliable predictions.

5.2 Rationale for Dataset Selection

The datasets were chosen for the following reasons:

1. Diversity in Features and Origin:

- The datasets include a wide range of clinical, demographic, and lifestyle attributes.
- They are sourced from different populations, enabling the evaluation of the generalizability of fuzzy ensemble models.

2. Class Imbalance:

- All three datasets exhibit varying degrees of class imbalance, providing an opportunity to test the effectiveness of SMOTE and other preprocessing techniques.

3. Structured Format:

- The datasets are structured and publicly available, making them suitable for reproducible research.

4. Real-World Relevance:

- The datasets reflect real-world scenarios in heart disease prediction, ensuring practical applicability of the research findings.

5.3 Preprocessing Steps

To prepare the datasets for analysis, the following preprocessing steps were performed:

5.3.1 Handling Missing Values

- Missing values were dropped rather than imputed, as imputation might compromise the uniqueness of patient-specific records.

5.3.2 Encoding Categorical Features

- Features such as "Sex," "Thalassemia," and "Smoking" were encoded using label encoding to ensure compatibility with machine learning algorithms.

5.3.3 Standardization

- Continuous features like "Age," "Cholesterol," and "BMI" were standardized to bring all attributes to a common scale, improving model performance and convergence.

5.3.4 Balancing Class Distribution

- SMOTE was applied to oversample the minority class in all three datasets, ensuring balanced training data and improving recall for the minority class.

5.4 Dataset Summary

The following table summarizes the key characteristics of the three datasets:

Table 5.1. Table summarizes the key characteristics of the three datasets.

Dataset Name	Description	Size	Features	Class Distribution
Heart Attack Prediction Dataset	Clinical, demographic, and lifestyle data for heart risk	8,763	14	Slightly imbalanced
Heart Attack Risk Prediction (Cleaned)	Refined data with lifestyle and health factors	9,651	15	Moderately imbalanced
Heart Attack Prediction in Indonesia	Region-specific patient data with diverse attributes	63,507	12	Highly imbalanced

5.5 Challenges in Datasets

The datasets pose several challenges that are addressed in the proposed work:

1. **Class Imbalance:**
 - All three datasets exhibit varying degrees of class imbalance, which can lead to biased predictions favouring the majority class. SMOTE is applied to mitigate this issue.
2. **Data Noise:**
 - The Indonesia dataset contains a higher degree of noise due to real-world inconsistencies, requiring robust models like Type-2 FRF and AFRF to handle uncertainty effectively.
3. **Feature Diversity:**
 - The datasets include a mix of categorical and continuous features, necessitating careful preprocessing and feature engineering.

CHAPTER 6

RESULTS AND DISCUSSION

6.1 Result

This section introduces five fuzzy logic-based models' classification performance, Type-1 Fuzzy RF, Type-2 Fuzzy RF, Adaptive Fuzzy RF, Weighted Fuzzy RF (FWRF), and Fuzzy Decision Forest (FDF)—compared on three datasets, both prior to and subsequent to SMOTE application. The models were measured with regular performance metrics: Accuracy, Precision, Recall, F1-Score, and AUC-ROC. The outcomes are tabulated in Tables for three datasets.

6.1.1 Performance on Heart Attack Prediction Dataset

Prior to the use of SMOTE, all models exhibited poor recall values indicating a strong weakness in identifying minority class instances as presented in Table 3. Though there were some instances of high precision, for instance, FDF achieved perfect precision (1.0000), models overall did not balance sensitivity and specificity. Type-2 Fuzzy RF model obtained the highest F1-score (0.1373) compared to all with low recall (0.0756), showing that it lacks strong robustness in dealing with class imbalance.

After applying Post-SMOTE, there were significant improvements in all the measures as can be seen in Table 4. The Type-2 Fuzzy RF model was the best performing with accuracy of 0.7693, precision of 0.8816, and highest F1-score (0.7373), and highest AUC-ROC value (0.8419) indicating superior discriminative power. Adaptive Fuzzy RF and Weighted Fuzzy RF also performed strongly with F1-scores of 0.7267 and 0.7299, respectively. The Type-1 Fuzzy RF had the best recall (0.6487), albeit with lower precision than Type-2 and Adaptive models. Altogether, results from Dataset 1 validate the effectiveness of SMOTE in boosting the minority class instance detection capability of fuzzy models without loss of precision.

Table 6.1. Evaluation Metrics for Fuzzy Logic-Based Models on Heart Attack Prediction Dataset Before Applying SMOTE.

Model	Accuracy	Precision	Recall	AUC-ROC	F1-Score
Type-1 Fuzzy RF	0.6546	0.0000	0.0000	0.5519	0.0000
Type-2 Fuzzy RF	0.6716	0.7424	0.0756	0.5761	0.1373
Adaptive Fuzzy RF	0.6711	0.7627	0.0694	0.5788	0.1273
Weighted Fuzzy RF (FWRF)	0.6764	0.8868	0.0725	0.5882	0.1341
Fuzzy Decision Forest	0.6572	1.0000	0.0077	0.5754	0.0153

Table 6.2. Evaluation Metrics for Fuzzy Logic-Based Models on Heart Attack Prediction Dataset After Applying SMOTE.

Model	Accuracy	Precision	Recall	AUC-ROC	F1-Score
Type-1 Fuzzy RF	0.6625	0.6772	0.6487	0.7358	0.6627
Type-2 Fuzzy RF	0.7693	0.8816	0.6335	0.8419	0.7373
Adaptive Fuzzy RF	0.7572	0.8552	0.6318	0.8324	0.7267
Weighted Fuzzy RF (FWRF)	0.7598	0.8578	0.6352	0.8284	0.7299
Fuzzy Decision Forest	0.7155	0.7651	0.6394	0.7827	0.6966

6.1.2 Performance on Heart Attack Risk Prediction (Cleaned)

The same trend was noticed in Heart Attack Risk Prediction (Cleaned). Prior to SMOTE, the models once more reported extremely low recall and F1-scores, reflecting a bad ability to find positive instances as presented in Table 5. For example, the Type-2 Fuzzy RF reported a recall of only 0.0143 and an F1-score of 0.0275, while the FDF and Type-1 Fuzzy RF models did not pick up any positive instances (recall = 0.0000). After applying SMOTE, the performance of all models was noticeably enhanced as indicated in Table 6. The highest accuracy (0.7590) and precision (0.8487) were achieved by the Type-2 Fuzzy RF, thus proving to be most accurate with regard to both correct classification and false positive reduction. The Adaptive Fuzzy RF recorded the highest AUC-ROC (0.8113) and F1-score of 0.7213, denoting a good balance of classification ability. Although the Type-1 Fuzzy RF had the highest recall of 0.6525, its precision and F1-score were relatively low. The FDF model was consistently moderate with an F1-score of 0.6657.

Table 6.3. Evaluation Metrics for Fuzzy Logic-Based Models on Heart Attack Risk Prediction (Cleaned) Before Applying SMOTE.

Model	Accuracy	Precision	Recall	AUC-ROC	F1-Score
Type-1 Fuzzy RF	0.6418	0.0000	0.0000	0.5000	0.0000
Type-2 Fuzzy RF	0.6366	0.3333	0.0143	0.5156	0.0275
Adaptive Fuzzy RF	0.6355	0.2105	0.0064	0.5078	0.0124
Weighted Fuzzy RF (FWRF)	0.6372	0.3571	0.0159	0.4905	0.0305
Fuzzy Decision Forest	0.6418	0.0000	0.0000	0.5109	0.0000

Table 6.4. Evaluation Metrics for Fuzzy Logic-Based Models on Heart Attack Risk Prediction (Cleaned) After Applying SMOTE.

Model	Accuracy	Precision	Recall	AUC-ROC	F1-Score
Type-1 Fuzzy RF	0.6445	0.6369	0.6525	0.7191	0.6447
Type-2 Fuzzy RF	0.7590	0.8487	0.6234	0.8095	0.7188
Adaptive Fuzzy RF	0.7562	0.8289	0.6384	0.8113	0.7213
Weighted Fuzzy RF (FWRF)	0.7417	0.8181	0.6139	0.7989	0.7015
Fuzzy Decision Forest	0.6845	0.6988	0.6356	0.7557	0.6657

6.1.3 Performance on Heart Attack Prediction in Indonesia

The performance of five fuzzy logic-based models on the Heart Attack Prediction in Indonesia dataset was evaluated using standard classification metrics both before and after the application of SMOTE, with results detailed in Tables 7 and 8. Initially, the models demonstrated relatively better performance on this dataset compared to the other two datasets, likely due to reduced class imbalance or a more uniform feature distribution. Type-2 Fuzzy RF emerged as the top-performing model in terms of F1-score (0.6472) and recall (0.6061), while FDF achieved the highest accuracy (0.7351) and precision (0.7116). Although Type-1 Fuzzy RF excelled in precision (0.7512), it had the lowest recall (0.5022), resulting in an F1-score of 0.6020. Following the application of SMOTE, all models exhibited notable improvements. Type-2 Fuzzy RF maintained its superior performance, achieving the highest F1-score (0.7888), recall (0.7999), and AUC-ROC (0.8787), indicating a well-balanced sensitivity and precision. Adaptive Fuzzy RF and FWRF closely followed with F1-scores of 0.7875 and 0.7871, respectively, demonstrating consistent classification capabilities. Type-1

Fuzzy RF and FDF also improved significantly, achieving F1-scores above 0.745 and demonstrating enhanced sensitivity to minority class instances.

Table 6.5. Evaluation Metrics for Fuzzy Logic-Based Models on Heart Attack Prediction in Indonesia Before Applying SMOTE.

Model	Accuracy	Precision	Recall	AUC-ROC	F1-Score
Type-1 Fuzzy RF	0.7326	0.7512	0.5022	0.8123	0.6020
Type-2 Fuzzy RF	0.7339	0.6943	0.6061	0.8072	0.6472
Adaptive Fuzzy RF	0.7270	0.6925	0.5793	0.8033	0.6308
Weighted Fuzzy RF (FWRF)	0.7303	0.6983	0.5814	0.8045	0.6345
Fuzzy Decision Forest	0.7351	0.7116	0.5754	0.8145	0.6363

Table 6.6. Evaluation Metrics for Fuzzy Logic-Based Models on Heart Attack Prediction in Indonesia After Applying SMOTE.

Model	Accuracy	Precision	Recall	AUC-ROC	F1-Score
Type-1 Fuzzy RF	0.7458	0.7438	0.7481	0.8371	0.7459
Type-2 Fuzzy RF	0.7864	0.7780	0.7999	0.8787	0.7888
Adaptive Fuzzy RF	0.7847	0.7754	0.7999	0.8768	0.7875
Weighted Fuzzy RF (FWRF)	0.7855	0.7792	0.7952	0.8777	0.7871
Fuzzy Decision Forest	0.7577	0.7559	0.7594	0.8489	0.7577

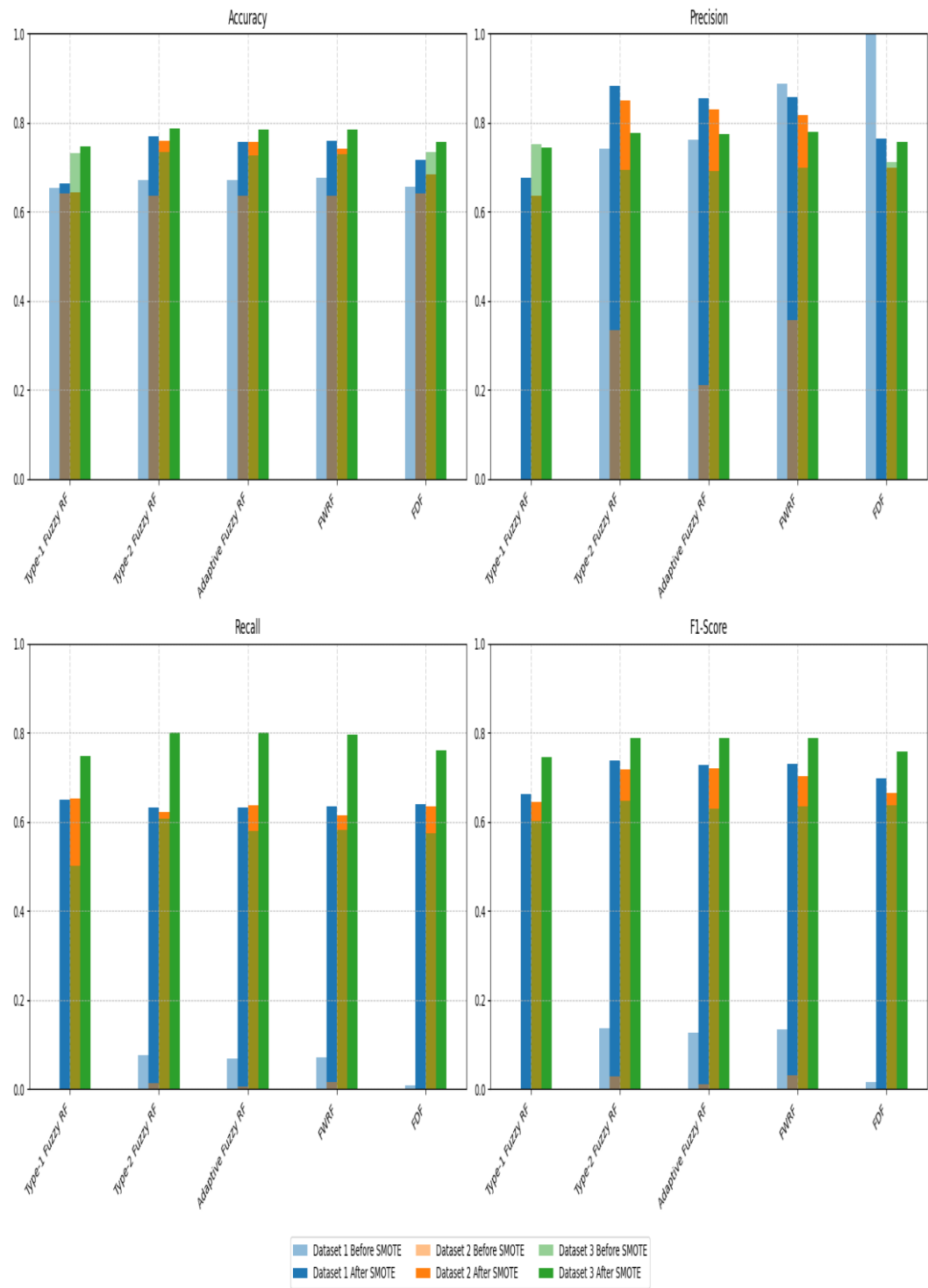


Fig. 6.1. Comparison of Accuracy, Precision, Recall, and F1-Score metrics before and after applying SMOTE across three datasets for five fuzzy logic-based models.

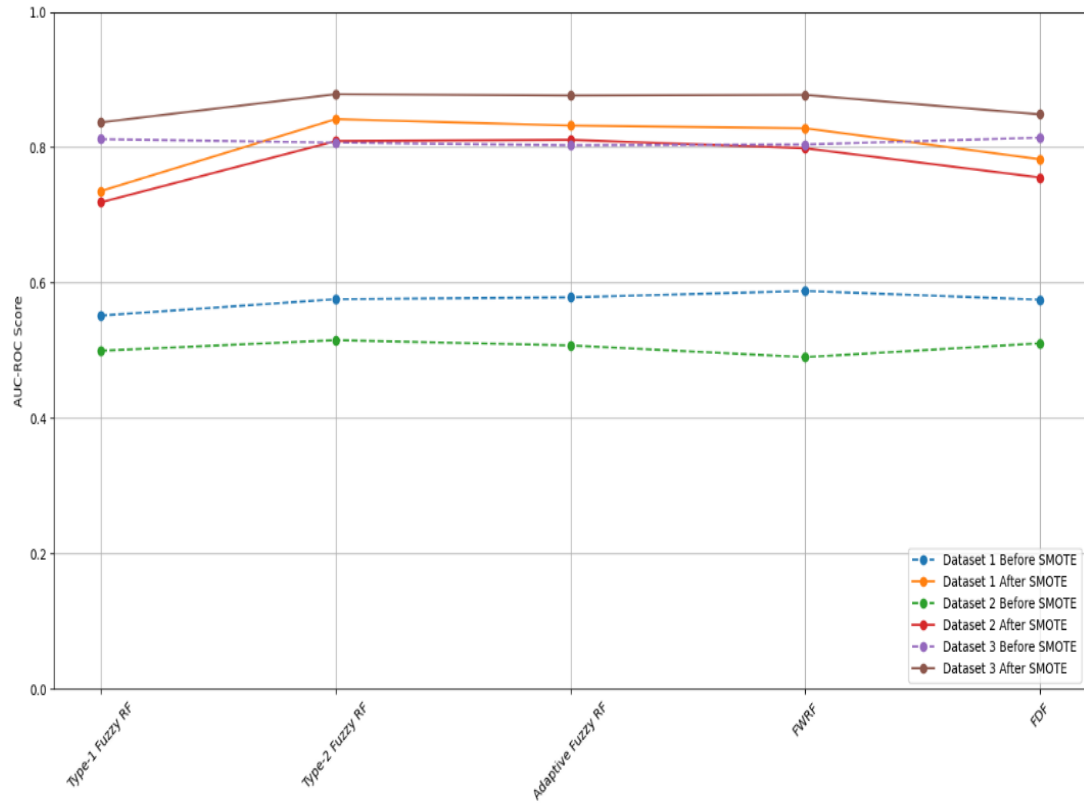


Fig. 6.2. AUC-ROC values of fuzzy logic-based models before and after SMOTE application across three datasets, illustrating improved discriminative performance.

6.2 Discussion

Comparative study of five ensemble classifiers constructed based on fuzzy logic, i.e., Type-1 Fuzzy RF, Type-2 Fuzzy RF, Adaptive Fuzzy RF, Weighted Fuzzy RF (FWRf), and Fuzzy Decision Forest (FDF), concluded the significant impact of class imbalance on model performance. Low recall and F1-scores were encountered in all three datasets prior to the application of SMOTE whereas relatively good precision was encountered. This imbalance heavily affected the models' performance in identifying minority class instances, an extremely critical requirement in mental illness detection issues. On using SMOTE, sharp improvement was seen in all performance metrics, most importantly recall, F1-score, and AUC-ROC. Type-2 Fuzzy RF and Adaptive Fuzzy RF always achieved the best precision-recall trade-off, indicating they are best applied to solve unbalanced classification. After using SMOTE techniques, Data Set 3 climbed to the top to achieve an F1-score of 0.7888 and an AUC-ROC of 0.8787 through Type-2 Fuzzy RF. It proves that applying SMOTE to the classifier improves its performance and sensitivity and that fuzzy-based models are important in healthcare complex classification.

CHAPTER 7

CONCLUSION AND FUTURE WORK

This thesis looks at the use of fuzzy logic-based approaches to predict heart disease in detail, with main attention on handling the problems linked to imbalanced datasets. From the research, we know that using fuzzy ensemble models and SMOTE boosts classification performance. The models examined were Type-1 Fuzzy Random Forest, Type-2 Fuzzy Random Forest, Fuzzy Weighted Random Forest (FWRF), Fuzzy Decision Forest (FDF) and Adaptive Fuzzy Random Forest (AFRF).

Results showed that Adaptive Fuzzy Random Forest always performed better than other models because of its smart membership function optimization. This technique also showed strong results, mainly because it handles uncertainty and noise in data well. Using SMOTE helped the model become more accurate and sensitive, addressing the usual problem of uneven class sizes in medical data.

Also, the study emphasizes that fuzzy logic is beneficial for managing unpredictable situations and giving understandable results essential for using it in healthcare. It shows that data preprocessing which includes standardization and changes to features, can lead to better results with machine learning models.

The study shows that fuzzy ensemble models can serve in diagnostics, particularly when predicting heart disease. This work uses fuzzy logic and SMOTE techniques to advance AI solutions for identifying patterns in healthcare datasets that have a high or low amount of data.

Future Work

While this research has achieved promising results, there are several areas for further exploration and improvement:

1. Integration with Deep Learning Models:

By mixing fuzzy ensemble models with CNNs or RNNs, we may improve the way complex medical data patterns are identified.

2. Real-World Clinical Validation:

Additional studies are needed where proposed models are used to analyze data from real patients, including both EHR and hospital records. By taking this step, the models can be applied more usefully in healthcare environments.

3. Hybrid Resampling Techniques:

Although SMOTE proved effective in this study, exploring advanced resampling methods, such as SMOTE-ENN or Borderline-SMOTE, may further enhance the quality of synthetic samples and improve model robustness.

4. Explainability and Interpretability:

Better interpretability of fuzzy ensemble models may make them more useful in healthcare. Integrating XAI methods such as XAI frameworks, can help explain what goes on when using machine learning models.

5. Multimodal Data Fusion:

By attempting to connect imaging, genomic and clinical data, future research can build diagnostic systems that draw on a broader range of input data.

6. Cost-Sensitive Learning:

Using cost-sensitive learning methods together with fuzzy logic might help achieve a fair balance between getting false positives and false negatives in important medical diagnosis situations.

By addressing these areas, the proposed models can be further refined to meet the demands of real-world healthcare applications, offering improved accuracy, robustness, and interpretability. These advancements would not only enhance heart disease prediction but also pave the way for broader applications of fuzzy logic in medical diagnostics.

References

- [1] R. Chitra, "Heart Attack Prediction System Using Fuzzy C Means Classifier," *IOSR J Comput Eng*, vol. 14, no. 2, pp. 23–31, 2013, doi: 10.9790/0661-1422331.
- [2] Haibo He and E. A. Garcia, "Learning from Imbalanced Data," *IEEE Trans Knowl Data Eng*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009, doi: 10.1109/TKDE.2008.239.
- [3] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, Jun. 2002, doi: 10.1613/jair.953.
- [4] A. K. Pathak and J. Arul Valan, "A Predictive Model for Heart Disease Diagnosis Using Fuzzy Logic and Decision Tree," 2020, pp. 131–140. doi: 10.1007/978-981-13-9680-9_10.
- [5] M. Pal and S. Parija, "Prediction of Heart Diseases using Random Forest," *J Phys Conf Ser*, vol. 1817, no. 1, p. 012009, Mar. 2021, doi: 10.1088/1742-6596/1817/1/012009.
- [6] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, Jun. 1965, doi: 10.1016/S0019-9958(65)90241-X.
- [7] A. Ebrahimzadeh and S. E. Mousavi, "Classification of communications signals using an advanced technique," *Appl Soft Comput*, vol. 11, no. 1, pp. 428–435, Jan. 2011, doi: 10.1016/j.asoc.2009.12.001.
- [8] K. Bahani, M. Moujabbir, and M. Ramdani, "An accurate fuzzy rule-based classification systems for heart disease diagnosis," *Sci Afr*, vol. 14, p. e01019, Nov. 2021, doi: 10.1016/j.sciaf.2021.e01019.
- [9] P. K. Anooj, "Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules," *Journal of King Saud University - Computer and Information Sciences*, vol. 24, no. 1, pp. 27–40, Jan. 2012, doi: 10.1016/j.jksuci.2011.09.002.
- [10] Ł. Gadomer and Z. A. Sosnowski, "Fuzzy Random Forest with C-Fuzzy Decision Trees," 2016, pp. 481–492. doi: 10.1007/978-3-319-45378-1_43.
- [11] Md. L. Ali, M. S. Sadi, and Md. O. Goni, "Diagnosis of heart diseases: A fuzzy-logic-based approach," *PLoS One*, vol. 19, no. 2, p. e0293112, Feb. 2024, doi: 10.1371/journal.pone.0293112.
- [12] M. G. El-Shafiey, A. Hagag, E.-S. A. El-Dahshan, and M. A. Ismail, "A hybrid GA and PSO optimized approach for heart-disease prediction based on random forest," *Multimed Tools Appl*, vol. 81, no. 13, pp. 18155–18179, May 2022, doi: 10.1007/s11042-022-12425-x.
- [13] G. T. Reddy and N. Khare, "Heart disease classification system using optimised fuzzy rule based algorithm," *Int J Biomed Eng Technol*, vol. 27, no. 3, p. 183, 2018, doi: 10.1504/IJBET.2018.094122.
- [14] B. K. Sarkar, "Hybrid model for prediction of heart disease," *Soft comput*, vol. 24, no. 3, pp. 1903–1925, Feb. 2020, doi: 10.1007/s00500-019-04022-2.
- [15] S. A. Mokeddem, "A fuzzy classification model for myocardial infarction risk assessment," *Applied Intelligence*, Dec. 2017, doi: 10.1007/s10489-017-1102-1.


- [16] U. Umoh, D. Asuquo, I. Eyoh, and V. Murugesan, “A comparative analysis of prediction problems utilizing Interval type-2 fuzzy and machine learning models,” *Int J Hybrid Intell Syst*, vol. 20, no. 4, pp. 301–316, Nov. 2024, doi: 10.3233/HIS-240008.
- [17] M. A. Jabbar, B. L. Deekshatulu, and P. Chandra, “Prediction of Heart Disease Using Random Forest and Feature Subset Selection,” 2016, pp. 187–196. doi: 10.1007/978-3-319-28031-8_16.
- [18] E. Zeinulla, K. Bekbayeva, and A. Yazici, “Effective diagnosis of heart disease imposed by incomplete data based on fuzzy random forest,” in *2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, IEEE, Jul. 2020, pp. 1–9. doi: 10.1109/FUZZ48607.2020.9177531.
- [19] S. H.-W. Chuah, E. C.-X. Aw, and D. Yee, “Unveiling the complexity of consumers’ intention to use service robots: An fsQCA approach,” *Comput Human Behav*, vol. 123, p. 106870, 2021, doi: <https://doi.org/10.1016/j.chb.2021.106870>.

LIST OF PUBLICATIONS

S.No.	Title of Paper	Conference Name	Status
1.	Machine Learning for Medical Diagnosis: A Review of Fuzzy Random Forests and Applications.	IEEE The Global AI Summit 2024.	Published
2.	Enhancing Fuzzy Logic-Based Classification of Imbalanced Datasets Using SMOTE: A Comparative Study Across Multiple Datasets.	6th International Conference on Data Analytics and Management (ICDAM-2025).	Presented

PUBLICATION PROOF

Acceptance Mail:


AVNEESH VERMA <avneeshverma05032000@gmail.com>

Global AI Summit 2024: Acceptance of Paper ID 252

1 message

Wed, Jul 24, 2024 at 2:50 PM

Microsoft CMT <email@msr-cmt.org>
 Reply-To: Jagendra Singh <jagendrasngh@gmail.com>
 To: Avneesh Verma <avneeshverma05032000@gmail.com>
 Cc: Mandeep.mittal@bennett.edu.in

Dear Author,

Greetings!

Thank you for submitting your research article to the 2024 IEEE The Global AI Summit 2024, to be held on Sep 04-06, 2024 at Bennett University, Greater Noida, India in Hybrid Mode.

We are pleased to inform you that based on reviewers' comments, your paper ID 252, title "Machine Learning for Medical Diagnosis: A Review of Fuzzy Random Forests and Applications" has been Provisionally accepted for presentation during The Global AI Summit 2024, and the conference proceedings to be published in IEEE Xplore, subject to the condition that you submit a revised version as per the comments, shared with the corresponding author. It is also required that you prepare a response to each comment from the reviewer and upload it as a separate file along with the revised paper. IEEE XPLORE is indexed with the world's leading Abstracting & Indexing (A&I) databases, including ISI / SCOPUS/ DBLP/ EI-Compendex / Google Scholar.

The similarity index in the final paper must be less than 20% including references. Please note that the high plagiarism, more than 0% AI-generated text and any kind of multiple submissions of this paper to other conferences or journals will lead to rejection at any stage.

Please Note that Conference proceedings that meet IEEE quality review standards will be eligible for inclusion in the IEEE Xplore Digital Library. IEEE reserves the right not to publish any proceedings or paper that do not meet these standards. Only presented paper will be included in proceedings.

Instruction for registration and camera-ready submission are available at:
<https://www.bennett.edu.in/aisummit2024/call-for-paper>
 Registration Link: <https://www.bennett.edu.in/aisummit2024/registration/>

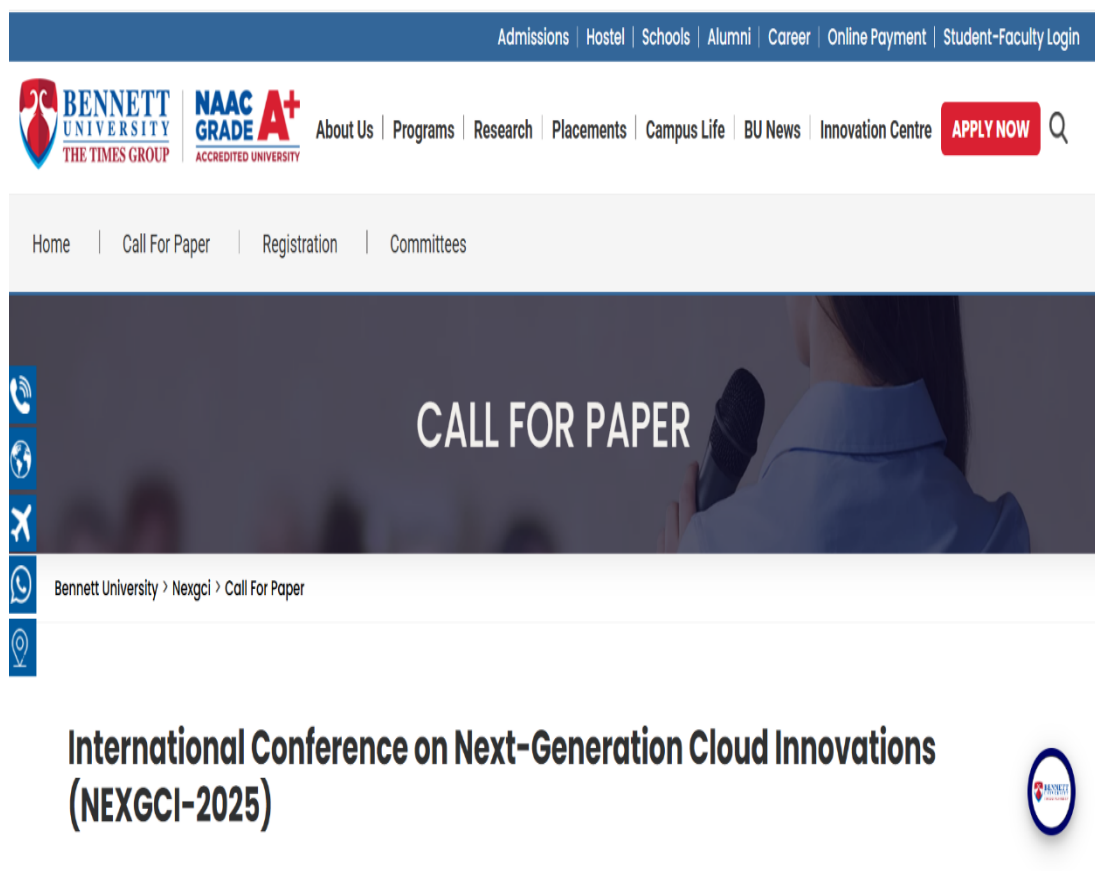
Please note that the Last date for submission of the camera-ready paper, payment of registration fee, and online registration is July 30, 2024.

Please remember to always include your Paper ID, whenever inquiring about your paper.


You can access review comments in your author console. Reviews are available under the status of the paper with link Reviews.

Looking forward to meeting you online during the conference.

With Regards
 Team Global AI Summit 2024
 AIML Track

Screenshot of website:

Indexing Proof:


AVNEESH VERMA <avneeshverma05032000@gmail.com>

Global AI Summit 2024: Acceptance of Paper ID 252

1 message

Wed, Jul 24, 2024 at 2:50 PM

Microsoft CMT <email@msr-cmt.org>
 Reply-To: Jagendra Singh <jagendrasngh@gmail.com>
 To: Avneesh Verma <avneeshverma05032000@gmail.com>
 Cc: Mandeep.mittal@bennett.edu.in

Dear Author,

Greetings!

Thank you for submitting your research article to the 2024 IEEE The Global AI Summit 2024, to be held on Sep 04-06, 2024 at Bennett University, Greater Noida, India in Hybrid Mode.

We are pleased to inform you that based on reviewers' comments, your paper ID 252, title "Machine Learning for Medical Diagnosis: A Review of Fuzzy Random Forests and Applications" has been Provisionally accepted for presentation during The Global AI Summit 2024, and the conference proceedings to be published in IEEE Xplore, subject to the condition that you submit a revised version as per the comments, shared with the corresponding author. It is also required that you prepare a response to each comment from the reviewer and upload it as a separate file along with the revised paper. IEEE XPLORE is indexed with the world's leading Abstracting & Indexing (A&I) databases, including ISI / SCOPUS/ DBLP/ EI-Compendex / Google Scholar.

The similarity index in the final paper must be less than 20% including references. Please note that the high plagiarism, more than 0% AI-generated text and any kind of multiple submissions of this paper to other conferences or journals will lead to rejection at any stage.

Please Note that Conference proceedings that meet IEEE quality review standards will be eligible for inclusion in the IEEE Xplore Digital Library. IEEE reserves the right not to publish any proceedings or paper that do not meet these standards. Only presented paper will be included in proceedings.

Instruction for registration and camera-ready submission are available at:
<https://www.bennett.edu.in/aisummit2024/call-for-paper>
 Registration Link: <https://www.bennett.edu.in/aisummit2024/registration/>

Please note that the Last date for submission of the camera-ready paper, payment of registration fee, and online registration is July 30, 2024.

Please remember to always include your Paper ID, whenever inquiring about your paper.

You can access review comments in your author console. Reviews are available under the status of the paper with link Reviews.


Looking forward to meeting you online during the conference.

With Regards
 Team Global AI Summit 2024
 AIML Track

Certificate of Participation:



Proof of Published:


Gmail

AVNEESH VERMA <avneeshverma05032000@gmail.com>

Call for Papers – 2nd Global AI Summit 2025 | IEEE Conference | Bennett University
1 message

Microsoft CMT <noreply@msr-cmt.org> Fri, May 9, 2025 at 10:06 AM
To: Avneesh Verma <avneeshverma05032000@gmail.com>

Dear Researcher/Academic Professional,

Greetings from Bennett University!

We are pleased to inform you that the AISUMMIT-2024 proceedings are now published on IEEE Xplore: <https://ieeexplore.ieee.org/xpl/conhome/10947709/proceeding>

We now invite you to submit your research to the 2nd Global AI Summit – International Conference on Artificial Intelligence and Emerging Technology (AI Summit 2025), scheduled for 19–21 November 2025 at Bennett University, Greater Noida (Hybrid Mode), in association with the University of Florida and registered with IEEE (Conference Record #66170).

Submission Deadline: 04 July 2025
Submit via CMT: <https://cmt3.research.microsoft.com/GlobalAISummit2025>
For Details, Please Visit: <https://www.bennett.edu.in/aisummit2025>

Thanks & Regards
Organizing Committee

To stop receiving conference emails, you can check the 'Do not send me conference email' box from your User Profile.

Microsoft respects your privacy. To learn more, please read our [Privacy Statement](#).

Microsoft Corporation
One Microsoft Way
Redmond, WA 98052

Conferences > 2024 International Conference... ?

Machine Learning for Medical Diagnosis: A Review of Fuzzy Random Forests and Applications

Publisher: IEEE

[Cite This](#)

[PDF](#)

Avneesh Verma ; Priyanka Arora ; Sonika Dahiya [All Authors](#)

11

Full

Text Views



Abstract

Document
Sections




Introduction

Abstract:

This paper explores Fuzzy Random Forests, a machine learning approach for classification tasks that combines ensemble classifiers, randomness, and fuzzy logic. Ensemble classifiers provide robustness, while randomness reduces correlation and increases diversity. Fuzzy logic allows the model to handle imperfect data by incorporating membership degrees. The research investigates Fuzzy

Acceptance Mail:


Gmail

AVNEESH VERMA <avneeshverma05032000@gmail.com>

ICDAM 2025: Paper Notification for Paper ID 1397
 2 messages

ICDAM Conf <icdam.conf@gmail.com>
 To: AVNEESH VERMA <avneeshverma05032000@gmail.com>

6th International Conference on Data Analytics & Management (ICDAM-2025)!

Dear Author(s),

Greetings from **ICDAM 2025!**

We congratulate you that your paper with submission ID **1397** and Paper Title '**Enhancing Fuzzy Logic-Based Classification of Imbalanced Datasets Using SMOTE: A Comparative Study Across Multiple Datasets**' has been accepted for publication in the Springer LNNS series [Indexing: SCOPUS, INSPEC, WTI Frankfurt eG, zbMATH, SCImago; All books published in the series are submitted for consideration in Web of Science]. This acceptance means your paper is among the top 20% of the papers received/reviewed. We urge you to complete your registration immediately to secure your spot at this highly anticipated event. **Please submit the revised paper by May 25th 2025, and complete the registration by May 25th, 2025 . No more further extension. Very few slots left.**

Please register as soon as possible and submit the following documents to icdam.conf@gmail.com as soon as possible.

1. Final Camera-Ready Copy (CRC) as per the springer format. (See <https://icdam-conf.com/downloads>)
2. Copy of e-receipt of registration fees. (For Registration, see <https://icdam-conf.com/registrations>)
3. The final revised copy of your paper should also be uploaded via Microsoft CMT.

Note : Standard Paper size – 10-12 pages. Over length of more than 12 pages, paper charges USD 20 per extra page

Tue, May 20, 2025 at 3:41 PM

Screenshot of website:



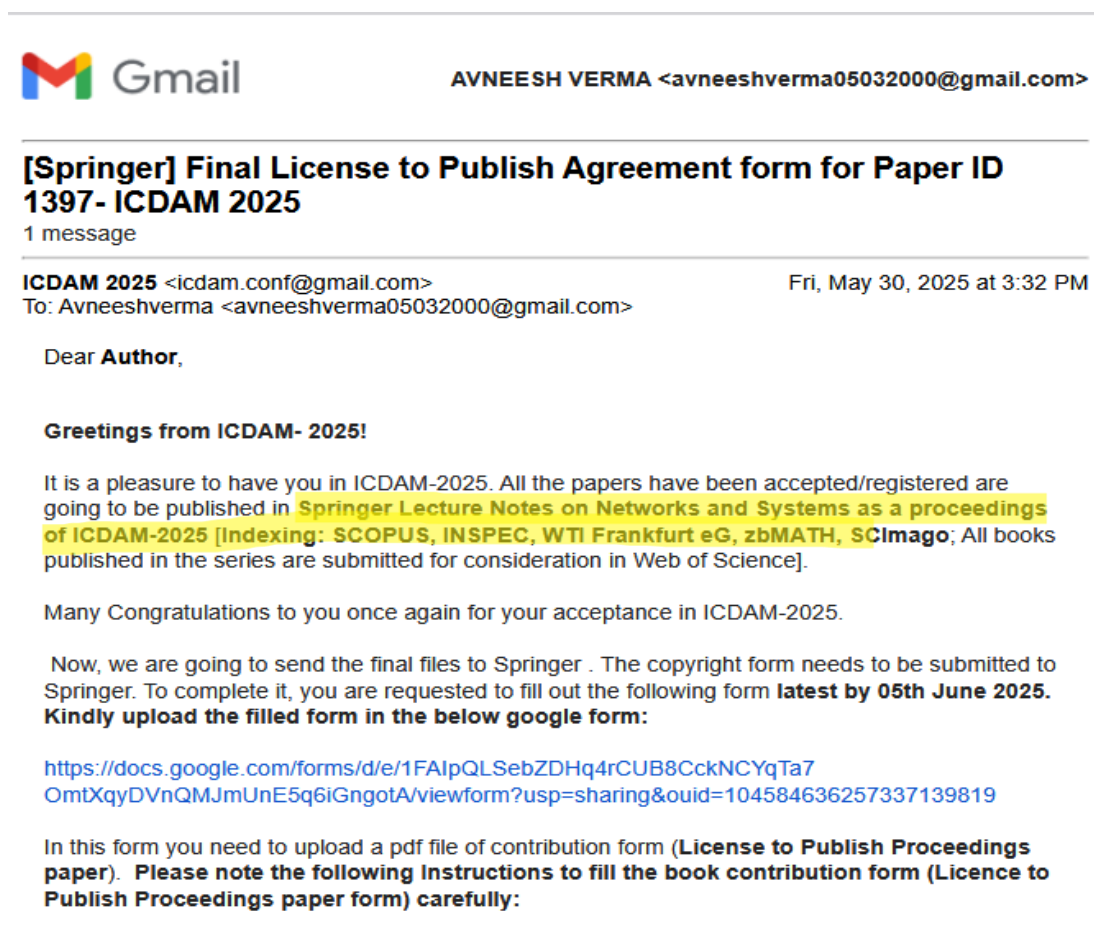
6th International Conference on Data Analytics & Management (ICDAM-2025)
ICDAM-2025 Theme: Data Analytics with Computer Networks

Organized By: London Metropolitan University, London, UK (Venue Partner)

In association with
WSG University, Bydgoszcz, Poland, Europe
&
Portalegre Polytechnic University, Portugal, Europe
&
SGGW Management Institute, Poland, Portugal

Date: 13th - 15th June, 2025
Springer LNNS Approved Conference (Indexed in Scopus, EI, WoS and Many More)

Indexing Proof:



[Springer] Final License to Publish Agreement form for Paper ID 1397- ICDAM 2025
 1 message

ICDAM 2025 <icdam.conf@gmail.com> Fri, May 30, 2025 at 3:32 PM
 To: Avneeshverma <avneeshverma05032000@gmail.com>

Dear **Author**,

Greetings from ICDAM- 2025!

It is a pleasure to have you in ICDAM-2025. All the papers have been accepted/registered are going to be published in **Springer Lecture Notes on Networks and Systems as a proceedings of ICDAM-2025 [Indexing: SCOPUS, INSPEC, WTI Frankfurt eG, zbMATH, SCImago]**; All books published in the series are submitted for consideration in Web of Science].

Many Congratulations to you once again for your acceptance in ICDAM-2025.

Now, we are going to send the final files to Springer . The copyright form needs to be submitted to Springer. To complete it, you are requested to fill out the following form **latest by 05th June 2025**. **Kindly upload the filled form in the below google form:**

<https://docs.google.com/forms/d/e/1FAIpQLSebZDHq4rCUB8CckNCYqTa7OmtXqyDVnQMjUnE5q6iGngotA/viewform?usp=sharing&ouid=104584636257337139819>

In this form you need to upload a pdf file of contribution form (**License to Publish Proceedings paper**). **Please note the following instructions to fill the book contribution form (Licence to Publish Proceedings paper form) carefully:**

Certificate of Participation:



PLAGIARISM ANNEXURE

Plagiarism Report:



Page 1 of 39 - Cover Page

Submission ID trn:old::27535:96900028

Avneesh Verma avneesh_Plag.docx

Delhi Technological University

Document Details

Submission ID
trn:old::27535:96900028

Submission Date
May 20, 2025, 11:40 PM GMT+5:30

Download Date
May 20, 2025, 11:43 PM GMT+5:30

File Name
avneesh_Plag.docx

File Size
662.7 KB

34 Pages
8,376 Words
48,625 Characters



Page 1 of 39 - Cover Page

Submission ID trn:old::27535:96900028





7% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.




Filtered from the Report

- Bibliography
- Quoted Text
- Cited Text
- Small Matches (less than 9 words)

Match Groups

-  **52 Not Cited or Quoted 7%**
Matches with neither in-text citation nor quotation marks
-  **0 Missing Quotations 0%**
Matches that are still very similar to source material
-  **0 Missing Citation 0%**
Matches that have quotation marks, but no in-text citation
-  **0 Cited and Quoted 0%**
Matches with in-text citation present, but no quotation marks

Top Sources

- 4%  Internet sources
- 2%  Publications
- 5%  Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

- 52 Not Cited or Quoted 7%**
Matches with neither in-text citation nor quotation marks
- 0 Missing Quotations 0%**
Matches that are still very similar to source material
- 0 Missing Citation 0%**
Matches that have quotation marks, but no in-text citation
- 0 Cited and Quoted 0%**
Matches with in-text citation present, but no quotation marks

Top Sources

- 4% Internet sources
- 2% Publications
- 5% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Internet	fastercapital.com	<1%
2	Internet	www.mdpi.com	<1%
3	Publication	Patel, Kaivan. "An Analytical Framework to Assess Liposuction Outcomes", Florid...	<1%
4	Submitted works	The University of Manchester on 2024-09-09	<1%
5	Submitted works	University of Stellenbosch, South Africa on 2024-08-04	<1%
6	Internet	citizenside.com	<1%
7	Submitted works		<1%
8	Publication	Abi, Beza Alemu. "Web SQL Injection Attack Detection Algorithm Using Deep Lear...	<1%
9	Internet	jurnal.polgan.ac.id	<1%
10	Publication	Al-Zaabi, Moza Sulaiman. "Predicting Students at Risk of Dropping Out in the Coll...	<1%

11	Submitted works	University of Greenwich on 2024-09-06	<1%
12	Internet	pubmed.ncbi.nlm.nih.gov	<1%
13	Internet	codeberg.org	<1%
14	Submitted works	University of East London on 2025-05-08	<1%
15	Internet	link.springer.com	<1%
16	Submitted works	Melbourne Institute of Technology on 2024-09-30	<1%
17	Submitted works	University of Oklahoma on 2020-11-30	<1%
18	Submitted works	City University of Hong Kong on 2025-04-30	<1%
19	Internet	www.frontiersin.org	<1%
20	Submitted works	University of Science and Technology, Yemen on 2015-03-21	<1%
21	Internet	www.nature.com	<1%
22	Internet	ebin.pub	<1%
23	Internet	fedetd.mis.nsysu.edu.tw	<1%
24	Internet	www.ejpam.com	<1%

25	Submitted works	Carnegie Mellon University on 2024-12-05	<1%
26	Submitted works	ICTS on 2025-05-06	<1%
27	Publication	S. Kaliraj, Veliseti Geetha Pavan Sahasranth, V. Sivakumar. "A holistic approach t...	<1%
28	Publication	Vishnu Vardhana Reddy Karna, Viswavardhan Reddy Karna, Varaprasad Janamala...	<1%
29	Internet	www.analyticsvidhya.com	<1%
30	Submitted works	Arts, Sciences & Technology University In Lebanon on 2023-12-27	<1%
31	Submitted works	IIT Delhi on 2012-09-05	<1%
32	Submitted works	Kampala International University on 2023-07-27	<1%
33	Submitted works	University of Northumbria at Newcastle on 2024-05-14	<1%
34	Publication	Yadav, Govind. "Enhancing the Accuracy of Large Language Models in Biomedical...	<1%
35	Submitted works	Icba on 2025-05-19	<1%
36	Internet	www.blog.qualitypointtech.com	<1%
37	Internet	www.fastercapital.com	<1%
38	Internet	www.geeksforgeeks.org	<1%

Avneesh Verma

avneesh_Plug.docx

 Delhi Technological University

Document Details

Submission ID**trn:oid::27535:96900028****Submission Date****May 20, 2025, 11:40 PM GMT+5:30****Download Date****May 20, 2025, 11:43 PM GMT+5:30****File Name****avneesh_Plug.docx****File Size****662.7 KB****34 Pages****8,376 Words****48,625 Characters**

*% detected as AI

AI detection includes the possibility of false positives. Although some text in this submission is likely AI generated, scores below the 20% threshold are not surfaced because they have a higher likelihood of false positives.

Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

Disclaimer

Our AI writing assessment is designed to help educators identify text that might be prepared by a generative AI tool. Our AI writing assessment may not always be accurate (it may misidentify writing that is likely AI generated as AI generated and AI paraphrased or likely AI generated and AI paraphrased writing as only AI generated) so it should not be used as the sole basis for adverse actions against a student. It takes further scrutiny and human judgment in conjunction with an organization's application of its specific academic policies to determine whether any academic misconduct has occurred.

Frequently Asked Questions

How should I interpret Turnitin's AI writing percentage and false positives?

The percentage shown in the AI writing report is the amount of qualifying text within the submission that Turnitin's AI writing detection model determines was either likely AI-generated text from a large-language model or likely AI-generated text that was likely revised using an AI-paraphrase tool or word spinner.

False positives (incorrectly flagging human-written text as AI-generated) are a possibility in AI models.

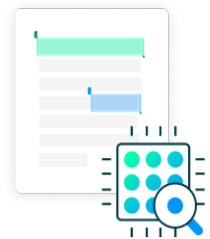
AI detection scores under 20%, which we do not surface in new reports, have a higher likelihood of false positives. To reduce the likelihood of misinterpretation, no score or highlights are attributed and are indicated with an asterisk in the report (*%).

The AI writing percentage should not be the sole basis to determine whether misconduct has occurred. The reviewer/instructor should use the percentage as a means to start a formative conversation with their student and/or use it to examine the submitted assignment in accordance with their school's policies.

What does 'qualifying text' mean?

Our model only processes qualifying text in the form of long-form writing. Long-form writing means individual sentences contained in paragraphs that make up a longer piece of written work, such as an essay, a dissertation, or an article, etc. Qualifying text that has been determined to be likely AI-generated will be highlighted in cyan in the submission, and likely AI-generated and then likely AI-paraphrased will be highlighted purple.

Non-qualifying text, such as bullet points, annotated bibliographies, etc., will not be processed and can create disparity between the submission highlights and the percentage shown.



DECLARATIONS

Declaration of Paper 1:

DECLARATION

We/I hereby certify that the work which is presented in the Major Project-II/Research Work entitled Analysis of Fuzzy Random Forest and its variants for breast cancer diagnosis in fulfillment of the requirement for the award of the Degree of Bachelor/Master of Technology in Software Engineering and submitted to the Department of Software Engineering, Delhi Technological University, Delhi is an authentic record of my/our own, carried out during a period from July-Dec-24, under the supervision of Dr. Sonika Dahiya.

The matter presented in this report/thesis has not been submitted by us/me for the award of any other degree of this or any other Institute/University. The work has been published/accepted/communicated in SCI/SCI expanded/SSCI/Scopus indexed journal OR peer reviewed Scopus indexed conference with the following details:

Title of the Paper: Machine Learning for Medical Diagnosis: A Review of Fuzzy Random Forests and Applications
 Author names (in sequence as per research paper): Auneesh Verma, Dr. Sonika Dahiya, Priyanka Anora
 Name of Conference/Journal: 2024 International Conference on Artificial Intelligence and Emerging Technology (Global AI Summit)
 Conference Dates with venue (if applicable): 14-17 September 24, Greater Noida
 Have you registered for the conference (Yes/No)? Yes
 Status of paper (Accepted/Published/Communicated): Accepted
 Date of paper communication: 20/07/2024
 Date of paper acceptance: 24/07/2024
 Date of paper publication: 09/04/2025

Auneesh
 Student(s) Roll No., Name and Signature
AUNEESH VERMA
23/SWE/15

SUPERVISOR CERTIFICATE

To the best of my knowledge, the above work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere. I, further certify that the publication and indexing information given by the students is correct.

Sonika
21/05/25
 Supervisor Name and Signature

Place: Delhi
 Date: 21-May-2025

NOTE: PLEASE ENCLOSE RESEARCH PAPER ACCEPTANCE/PUBLICATION/COMMUNICATION PROOF ALONG WITH SCOPUS INDEXING PROOF (Conference Website OR Science Direct in case of Journal Publication).

Declaration of Paper 2:

DECLARATION

We/I hereby certify that the work which is presented in the Major Project-II/Research Work entitled Analysis of Fuzzy Random Forest And Its Variants for heart attack prediction in fulfillment of the requirement for the award of the Degree of Bachelor/Master of Technology in Software Engineering and submitted to the Department of Software Engineering, Delhi Technological University, Delhi is an authentic record of my/our own, carried out during a period from Jan to May 25, under the supervision of Dr. Sonika Dahiya.

The matter presented in this report/thesis has not been submitted by us/me for the award of any other degree of this or any other Institute/University. The work has been published/accepted/communicated in SCI/SCI expanded/SSCI/Scopus indexed journal OR peer reviewed Scopus indexed conference with the following details:

Title of the Paper: Enhancing Fuzzy Logic-Based Classification of Imbalanced Datasets using SMOTE: A Comparative Study Across Multiple Datasets
 Author names (in sequence as per research paper): Aneesh Verma, Dr. Sonika Dahiya
 Name of Conference/Journal: TEDAM-2025
 Conference Dates with venue (if applicable):
 Have you registered for the conference (Yes/No)? Yes
 Status of paper (Accepted/Published/Communicated): Accepted
 Date of paper communication: 17/May/25
 Date of paper acceptance: 20/May/25
 Date of paper publication:

Aneesh
 Student(s) Roll No., Name and Signature
Aneesh Verma
23/SWE/15

SUPERVISOR CERTIFICATE

To the best of my knowledge, the above work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere. I, further certify that the publication and indexing information given by the students is correct.

Sonika
21/05/25
 Supervisor Name and Signature

Place: Delhi
 Date: 21-May-2025

NOTE: PLEASE ENCLOSE RESEARCH PAPER ACCEPTANCE/PUBLICATION/COMMUNICATION PROOF ALONG WITH SCOPUS INDEXING PROOF (Conference Website OR Science Direct in case of Journal Publication).