International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 16 (2024) pp. 515–534

Article

Fuzzy Logic and Machine Learning Integration: Enhancing Healthcare Decision-Making

Kirti Gupta [1,*], Pardeep Kumar [1], Shuchita Upadhyaya [1], Monika Poriye [1] and Shalini Aggarwal [2]

[1] Department of Computer Science & Applications, Kurukshetra University, Kurukshetra 136119, India; mittalkuk@gmail.com (P.K.); shuchita_bhasin@yahoo.com (S.U.); monikaporiye@gmail.com (M.P.)
[2] Department of Computer Science, Shaheed Udham Singh Govt College, Karnal 132041, India; shaliniagg07@gmail.com
* Correspondence author: kirtigupta@kuk.ac.in

Received date: 1 March 2024; Accepted date: 21 March 2024; Published online: 10 July 2024

Abstract: The escalating healthcare costs in contemporary society have raised significant concerns. Identifying medical risks efficiently is crucial for reducing treatment expenses and improving overall health outcomes. However, the current disease risk assessment process involves multiple tests and requires medical professionals' expertise, leading to time-consuming and expensive procedures. In response to these challenges, the current role of machine learning in healthcare holds promise by offering efficient solutions for disease risk assessment, potentially streamlining processes, and contributing to cost reduction while improving health outcomes. However, the classification of medical diseases using machine learning (ML) algorithms presents challenges due to the presence of incomplete, uncertain, and inaccurate data. This research paper conducts a comprehensive survey of prior studies in the application of ML techniques for disease diagnosis, emphasizing the need for a system capable of integrating both linguistic and numeric inputs to enhance the diagnostic process's robustness. The study aims to surpass merely improving clinical outcomes by focusing on enhancing diagnostic accuracy, optimizing patient care, and resource utilization. It further explores machine learning (ML) techniques for disease diagnosis, introducing a hybrid ML-fuzzy logic (FL) model evaluated on five healthcare datasets related to diabetes, heart stroke, heart failure, and body fat predictions. The empirical findings and evaluations are conducted using the Python 3.8.3 environment with Jupyter Notebook. Seven existing ML algorithms, alongside the proposed hybrid Fuzzy-PCA-SVM model, are employed on all datasets. To evaluate the model's effectiveness, various performance standards, including accuracy, precision, F1-score, and recall, have been considered. The results demonstrate that by leveraging the benefits of both SVM and FL systems, the suggested hybrid model outperforms other ML models. The study not only underscores the significance of integrating linguistic and numeric inputs in disease diagnosis but also envisions future research focused on real-world datasets and improved feature selection techniques for continued advancements in healthcare analytics.

Keywords: artificial intelligence (AI); fuzzy logic (FL); machine learning (ML); support vector machines (SVM); random forest (RF); decision tree (DT); Naïve Bayes (NB)

1. Introduction

In the landscape of digital healthcare, a multitude of opportunities arises for continuous data tracking, enhanced clinical outcomes, and the reduction of human errors. Artificial Intelligence (AI) techniques, ranging from ML to deep learning, play pivotal roles in various well-being-related domains. These include the treatment of diverse illnesses, the development of advanced clinical systems, and the management of patient data and records [1–4]. Notably, ML and deep learning applications prove instrumental in identifying and studying illnesses that pose challenges for conventional diagnostic methods. In [5], a comprehensive analysis employing AI methodologies for diagnosing a spectrum of challenging illnesses, including diabetes, chronic heart disease, TB, stroke, hypertension, skin, and liver diseases has been carried out.

Within healthcare decision-making, information often exhibits uncertainties, presenting challenges for technology producers and researchers. Traditional logic tends to categorize information into binary outcomes, yet FL provides a nuanced framework that captures the “grey areas” in between [6]. While integrating ML techniques with FL has shown promise in healthcare classification tasks, Refs. [7,8] the development of high-performance training methods remains a challenge.

This study aims to address this gap by proposing a model capable of handling both linguistic and numeric inputs, catering to non-specialists in healthcare. The primary objectives are to enhance diagnoses, improve patient care, and optimize resource utilization. Given the advancements in AI and ML, particularly in disease prevention, personalized medicine, and digital diagnosis, there is a pressing need for high-quality ML and AI decision support systems to effectively address healthcare challenges and elevate patient outcomes.

This research paper explores the application of FL systems in healthcare, highlighting their prowess in manipulating and representing uncertain information. By leveraging FL, medical professionals gain a valuable tool to navigate linguistic nuances and mitigate the loss of clarity in decision-making processes. The effectiveness of FL technologies, recognized across the medical field, demonstrates their capability to address the inherent fuzziness associated with healthcare scenarios.

The amalgamation of FL with ML techniques enables the synergistic leveraging of their strengths. ML algorithms, adept at handling vast amounts of data and uncovering patterns, enhance the usefulness and efficiency of the FL framework. This paper underscores the significance of integrating FL with ML, paving the way for the development of advanced methodologies that provide essential support to medical professionals, ultimately leading to improved patient care outcomes. The subsequent sections provide a detailed discussion of ML, FL, healthcare, and the pivotal role played by the integration of ML and FL in healthcare processes.

1.1. Machine Learning (ML) and Its Applications in Healthcare

ML is a subset of AI, that encompasses the process of analyzing data to enhance performance in various tasks. ML is a subset of artificial intelligence (AI) that involves the development of algorithms and statistical models to enable computer systems to learn and improve their performance without explicit programming. ML algorithms analyze data, recognize patterns, and make informed decisions, ultimately enhancing their ability to handle complex tasks [9]. The integration of computational statistics and statistical learning enhances the capabilities of ML, with contributions from mathematical optimization tools, theoretical frameworks, and practical applications. Data scientists employ exploratory data analysis (EDA) and data visualization techniques for effective dataset summarization and analysis.

In emulation of the intricate workings of the human brain, certain ML methodologies employ neural networks and data-driven models to mirror cognitive functions. Renowned in business realms as predictive analytics, ML adeptly leverages historical data to make precise predictions [10]. This notable advancement extends to the healthcare sector, where ML utilization has experienced significant growth. Through the training of algorithms on extensive healthcare datasets, ML holds the promise of providing accurate predictions and valuable insights, thereby enhancing decision-making processes within the healthcare domain.

ML in healthcare offers diverse applications:

- **Disease Diagnosis and Prediction:** ML models analyze medical data for diseases like cancer and diabetes, enabling early detection and personalized treatment plans [11].
- **Drug Discovery and Development:** ML accelerates drug discovery by predicting drug candidates and optimizing clinical trials.
- **Personalized Medicine:** ML tailors treatments based on individual characteristics, enhancing efficacy and minimizing adverse effects.
- **Healthcare Management:** ML optimizes hospital operations, predicts patient admission rates, and aids in managing healthcare supply chains.
- **Image and Speech Recognition:** ML interprets medical images for accurate diagnosis and transcribes medical notes through speech recognition.
- **Remote Patient Monitoring:** ML supports real-time health insights through analyzing data from wearable devices, aiding in early intervention.

- **Fraud Detection and Security:** ML enhances healthcare fraud detection and utilizes biometric authentication for data security.
- **Natural Language Processing (NLP):** NLP improves electronic health record management and facilitates information retrieval in healthcare.

This research paper delves into the Proliferating role of ML in healthcare, emphasizing its ability to enhance predictive analytics and support informed decision-making. By harnessing ML's capabilities, healthcare practitioners can extract valuable insights from historical data, leading to more precise diagnoses, improved treatment planning, and enhanced patient care [6]. The fundamental principles of ML methodologies and their applications in healthcare are elucidated in Table 1.

1.2. Fuzzy Logic in Healthcare Decision-Making

Fuzzy logic (FL), introduced by Lotfi Zadeh in 1965, offers a distinctive mathematical approach that accommodates intermediate values between true and false, effectively handling imprecise or uncertain data in decision-making processes. Unlike traditional binary logic that relies on rigid distinctions, such as yes/no or true/false, FL embraces a framework that captures nuanced "grey areas" that exist in between. This flexibility makes it particularly well-suited for applications in healthcare, where information often exhibits uncertainty and requires a more nuanced interpretation.

In healthcare scenarios, where imprecision and uncertainty are inherent, FL provides a valuable tool to manage linguistic notions and mitigate the resulting loss of clarity in decision-making processes. This proven effectiveness has led to the widespread recognition and utilization of FL technologies in addressing the complexities associated with healthcare scenarios.

The following sections will delve into a comprehensive exploration of FL, its applications in healthcare, and its integration with ML techniques to enhance decision-making processes in this critical domain.

1.3. Exploring the Transformative Potential of Integrating ML and FL in Healthcare Decision-Making

The motivation behind this research stems from the profound potential of synergizing ML and FL in healthcare, presenting a transformative paradigm capable of addressing the complexities and uncertainties inherent in healthcare scenarios. The adaptive learning capabilities of ML algorithms, coupled with the nuanced reasoning of FL, form a hybrid model that excels in enhancing diagnostic precision, patient care outcomes, and the overall efficiency of healthcare resource allocation.

Specifically, the research is driven by the recognition that the integration of ML and FL offers enhanced disease diagnosis precision. Leveraging extensive healthcare datasets, the ML component discerns intricate patterns, while FL interprets linguistic aspects of medical knowledge, contributing to a more refined diagnostic process by handling imprecise or uncertain information.

The work is further motivated by the potential for improved patient care outcomes through the adaptive learning of ML algorithms. These algorithms facilitate personalized treatment plans based on individual patient characteristics and responses to therapies, complemented by fuzzy logic's capacity to consider partial truths and uncertainties, tailoring medical interventions to the diverse and dynamic nature of patient conditions.

The purpose of this study is to develop an advanced model that integrates ML techniques, specifically in conjunction with FL, to address the challenges in healthcare classification tasks. The model should be capable of effectively handling both linguistic and numeric inputs, catering to non-specialists in the field. The main goals of this research are to enhance the accuracy of diagnoses, improve patient care, and optimize resource utilization in the healthcare domain.

The subsequent sections of the paper are structured as follows: Section 2 presents a concise review of relevant literature, highlighting key insights drawn from existing research. Section 3 outlines the suggested methodology, encompassing data acquisition and the foundational framework of the proposed approach. In Section 4, the Experimental Evaluation of Hybrid Model Integration in Healthcare Classification is elaborated, providing details on the experimental design and outcomes. Section 5 engages in a thorough discussion of the results obtained. The paper concludes in Section 6 with a summary of findings, insights, and avenues for future research.

Table 1. Comprehensive Classification of ML Methodologies in Healthcare.

Methodology	Description	Application
Supervised Learning	Utilizes labelled training data to train the model, allowing it to make predictions or decisions based on input features.	Medical diagnosis, treatment planning [12]
Unsupervised Learning	Learns patterns and relationships within data without labelled examples. Useful for identifying hidden structures and trends.	Patient clustering, anomaly detection [13]
Reinforcement Learning	Involves training models to make sequences of decisions. The model learns from trial and error, receiving feedback in the form of rewards or penalties.	Treatment optimization, personalized medicine [14,15]
Deep Learning	A subset of ML involving neural networks with multiple layers (deep neural networks). Enables complex pattern recognition and feature extraction.	Medical imaging analysis, pathology detection [16,17]
Ensemble Learning	Combines multiple models to enhance overall performance. Ensemble methods include bagging, boosting, and stacking.	Predictive modeling, improving model robustness [12,18]
Transfer Learning	Involves training a model on one task and applying the knowledge gained to a different but related task.	Leveraging pre-trained models for specific healthcare tasks [19,20]
Semi-Supervised Learning	Uses a combination of labeled and unlabeled data for training. Useful when acquiring labeled data is challenging or expensive.	Limited labeled datasets, data with mixed annotations [21,22]
Meta-Learning	Focuses on learning how to learn. Aims to enable models to adapt quickly to new tasks with minimal data.	Rapid adaptation to emerging healthcare challenges [23,24]
Online Learning	Involves continuous learning from new data as it becomes available. Suitable for scenarios where data streams are constant and evolving.	Real-time monitoring and adaptation in healthcare systems [25]
Explainable AI (XAI)	Emphasizes making ML models understandable and interpretable. Essential for ensuring transparency and trust in healthcare decision-making.	Enhancing interpretability of diagnoses and decisions [26]

2. Literature Review

A literature review is the foundation of any study. As a result, a substantial literature review on ML and FL in healthcare was conducted. Some of the significant work is provided below:

The ML techniques that are used in the medical and health sciences have increased significant attention in recent years. Researchers have conducted a comprehensive assessment of ML techniques and their application in creating reliable and adaptable prediction models [5,27]. This underscores the significance of continuously improving and advancing ML techniques for accurate and effective prediction models in the medical field. Other studies, like those for applications like disease prediction and medical diagnosis use FL, clustering, and SVM methods [6,28]. These studies demonstrate the potential of ML in addressing various challenges in healthcare, including improving prediction accuracy, reducing network latency, enhancing diagnostic services, and optimizing green computing processes. As technology continues to evolve, the integration of AI, ML, and data mining methods in healthcare holds great promise for transforming medical practices and improving patient outcomes.

The tabularized review of the literature, as shown in Table 2, offers a thorough overview of the many research projects in the fields of medical and health sciences. The summaries and key inferences are presented for each study, highlighting the main findings and contributions.

Table 2. Advancements and Applications of ML, FL, and AI in Medical and Health Sciences: A Comprehensive Review.

Authors	Summary	Inference
Alanazi et al. [28]	A comprehensive evaluation of the utilization of ML methods for developing reliable and adaptable predictive models within the realm of medical and health sciences is conducted. Our study scrutinized advanced healthcare	Different predictive algorithms yield varied results and proposed models require further refinement.

	prediction models, aiming to underscore the diversity of outcomes generated by different predictive algorithms and the imperative for ongoing enhancement in this domain.	
Kushwaha et al. [29]	Introduced a fuzzy magnetic optimization clustering approach and its utilization in the field of medicine. Explored the significance of clustering in data mining and knowledge discovery to unveil latent patterns in massive datasets.	Clustering aids in identifying patterns and clusters in large medical datasets.
Uddin et al. [30]	Compared various supervised algorithms of ML for the prediction of illness. The authors emphasized the significance of supervised ML techniques in data mining and the promise of leveraging health data to predict disease.	Supervised ML algorithms show promise in disease prediction using health data.
Graham et al. [31]	The authors examined the use of AI in mental health and offered a summary of recent studies focusing on the utilization of AI in mental health, emphasizing its capacity to complement clinical approaches. The paper also explored its potential benefits alongside acknowledging constraints, areas requiring further research, and ethical considerations.	AI can complement clinical practice in mental health but additional research and ethical considerations are necessary.
Shukla et al. [32]	Investigated a three-tier architecture to reduce network latency in healthcare IoT they used fog computing and ML. Proposed a hybrid approach combining FL and reinforcement learning for healthcare IoT analysis, aiming to improve connection speed.	The integration of FL and reinforcement learning within fog computing is geared towards minimizing network latency in the context of healthcare IoT.
Bhatt et al. [33]	Conducted a survey on chronic kidney disease diagnosis using FL. Explored medical expert systems designed for illness diagnosis using various methods, focusing on the rising prevalence of chronic kidney disease.	FL is employed for diagnosing chronic kidney disease, a rising global health concern.
Singh Tomar et al. [8]	Reviewed a Python-based fuzzy classifier for cashew kernels. Discussed the application of FL, specifically in the Python language, for implementing a simple fuzzy classifier and compared it to traditional FL tools.	Python libraries offer an effective alternative to traditional FL tools in implementing fuzzy classifiers.
Zahra Benchara et al. [34]	Introduced a decentralized methodology employing a massively parallel and distributed virtual mobile agent architecture for the analysis of MRI data, showcasing the accuracy and effectiveness of the proposed method. The approach presented a distributed type-2 FL method designed to enhance the efficiency of medical informatics data science models.	The enhancement of medical data analysis is achieved through the integrated architecture.
Casalino et al. [35]	Introduced a hierarchical fuzzy system for assessing the risk of cardiovascular disease. Proposed a hierarchical fuzzy inference system (HFIS) to predict the severity of cardiovascular disease, improving classification performance and interpretability compared to simple fuzzy inference systems.	A hierarchical fuzzy system improves classification performance and interpretability in predicting cardiovascular disease severity.
Omoregbe et al. [36]	Developed a text message-based medical diagnostic service using NLP and FL. Explored the application of NLP methods in conversational systems for health diagnosis, improving the accuracy of medical diagnoses based on user input.	NLP-based conversational systems improve medical diagnosis and provide accurate prognoses through user input.

Zubar et al.[37]	Examined the optimization of green computing using ML algorithms in healthcare. Explored the use of hybrid optimization methods in big data analytics to address communication issues in healthcare networks, with a focus on heart disease and related topics.	Data mining and optimization methods enhance decision-making in healthcare networks and address communication challenges.
Reddy et al. [38]	Presented a hybrid genetic algorithm and FL classifier for the detection of cardiovascular illness.	Early detection and treatment of cardiovascular illness can mitigate mortality rates and medical expenses.
Jayalakshmi et al. [39]	Health monitoring systems for COVID-19 patients often involve the collection and analysis of various physiological parameters such as temperature, oxygen saturation, respiratory rate, and more. The authors proposed a FL system to track COVID-19 patient's health.	Three distinct classification models are employed to recognize patient activities and medical history. Among these, the fuzzy adapted model demonstrates the highest accuracy.
Ullah et al. [7]	The Type-1 membership function, characterized by a single membership value, inadequately represents the uncertainty in the degree of membership. Consequently, when there is substantial uncertainty in both the source and fused data, the model's reliability becomes questionable. Decision-making using Type-2 FL (T2FL) has been the subject of recent research. T2FL, a FL generalization, can handle higher-order types of uncertainty in the source data. To address the uncertainty present in the decision-making system's input, a novel scheme for successfully integrating T2FL with Dempster–Shafer evidence theory (DST) is put forth in this paper.	The suggested scheme performs significantly better in terms of inference accuracy than the current schemes based on ontology and type-1 FL, on the datasets for diabetes and heart disease.
Harb, H. et al. [40]	Incorporating efficient data analytics based on sensed data to support medical and hospital staff by monitoring and assessing patients in real time was done in [19]. They used the Long Short-Term Memory (LSTM) model in their paper.	The desired level of accuracy required to predict the course of the patient's condition heavily influences the processing complexity of the LSTM; more processing complexity is required to increase the accuracy of the LSTM, and vice versa.
Duggento, A. et al. [41]	In the paper, abnormal phonocardiograms are detected using CNN and the Mel-frequency spectrum.	Further efforts are needed to fine-tune the classification architecture, specifically focusing on heart-phase-based classification, to achieve a more refined and precise clinical diagnosis.
Mostafa et al. [42], Annamalai et. al. [43]	FL-based AI analysis has been applied to the suggestion, development, and testing of Internet of Medical Things (IoMT)-based remote health monitoring.	This paper proposed a simple-to-use device that analyzes recorded sensor values using the FL system available in Arduino libraries to directly provide the risk ratio.
Hameed et al. [44]	In an emergency, patients could benefit from remote monitoring thanks to a proposed IoT-based healthcare infrastructure.	The system made intelligent decisions for patient care, management, and monitoring by using neural networks and FL systems to identify potential conditions and treatments.

Alshamrani et al. [45], Rahman et al [46], Sharma [47]	Some works describe an IoT-based framework for monitoring remote patients, such as the ones presented [25–27]. These papers describe an architecture for an Internet of Things system based on three-layer modeling: a hardware module with a few sensors for vital signs, and a gateway layer that gathers, stores, and makes data available for the application layer, which is the higher level.	Although the use of various protocols varies in the communication layer, overall, they adhere to a similar architecture, which is becoming the norm in many new works such as this one.
Tigga et al. [48]	The goal of the authors was to evaluate the patient's risk of diabetes based on their daily activities, how they live life, health issues, etc. They conducted experiments on 952 records, which were gathered through an online and offline survey. The Pima Indian Diabetes database was treated in the same way.	The Random Forest (RF) classifier demonstrates superior performance in evaluating the risk of diabetes among patients, indicating its effectiveness in medical prediction tasks.

The inferences drawn in table highlight the diverse applications and potential of ML, FL, and AI techniques in addressing healthcare challenges and improving patient outcomes.

The integration of ML algorithms and FL systems has been extensively explored in recent studies [49]. Researchers have developed theoretically sound models to facilitate well-informed decision-making and forecast patients' satisfaction with telemedicine [50]. Additionally, simulations for heart disease prediction using the partly observable Markov decision process (POMDP) have been proposed [51], with detailed design, implementation, testing, and analysis presented [52]. The Inhaler Compliance Assessment tool has been utilized to record patients using a Diskus dry powder inhaler [53], while a scalable system capable of detecting falls, tracking thousands of senior citizens, and notifying caregivers has been demonstrated [54]. ML has also been employed to determine a patient's medical specialty based on their symptoms [55]. Furthermore, research has focused on the advantages of various medical imaging applications [46], with a threshold approach utilized to automatically split breath cycles into smaller cycles [56].

The application of various techniques, including SVM, KNN, DT, Ada Boost, RF, K-Mean clustering, RNN, CNN, Deep-CNN, FL, LSTM, and others, for disease detection systems is essential for understanding how AI aids in disease diagnosis and prediction [57–59]. Observational studies employing ML prediction techniques have been used to predict diseases in various contexts [60,61]. ML techniques are increasingly being utilized to diagnose and predict a wide range of diseases, including cardiovascular diseases, cancers, diabetes, hepatitis, and tuberculosis (TB).

These studies highlight the diverse applications of ML techniques, FL, and AI in different aspects of medical and healthcare domains. They demonstrate the potential for improving prediction models, disease diagnosis, data analysis, IoT systems, and patient care through the integration of advanced technologies. Despite the potential benefits, the majority of hospitals do not currently use ML technologies, largely due to challenges such as lack of expertise among healthcare professionals in creating and deploying proficient models. In response to these challenges, a growing field is dedicated to automatically selecting, composing, and parameterizing ML and FL models.

3. Method

This section provides a foundation for the proposed work with an exploration and comparison of various ML techniques, with a focus on identifying their strengths and limitations to optimize performance for specific tasks or datasets, aiming to reduce reliance on human experts. In response to challenges in performance, scalability, and adaptability, this paper proposes a hybrid approach that combines FL with ML. The objective is to develop a hybrid model. Tailored to healthcare, addressing the intricacies of datasets, steps for model generation, and the foundational aspects of the proposed framework across three subsections.

3.1. Datasets

The study utilizes five datasets namely the Pima Indian Diabetes Dataset, Heart Stroke prediction dataset, Heart Failure Prediction Dataset, Body Fat Prediction Dataset, and Heart Disease Dataset Comprehensive outlined in Table 3. The Source of the datasets, the number of Attributes, and the records in each dataset are elaborated in Table 3. These datasets undergo initial cleaning and preprocessing procedures to eliminate missing and null values. Then these datasets were used for carrying out the data pre-processing steps. During the feature extraction phase, correlations between the features are identified. Subsequently, different ML algorithms are employed, leveraging the selected features, to predict the health status of individuals.

Table 3. Dataset Used.

Datasets	Website	No. of Attributes	No. of Instances
Pima Indian Diabetes Dtaset	Kaggle	9	768
Heart Stroke Prediction Dataset	Kaggle	12	5110
Heart Failure Prediction Dataset	Kaggle	11	918
Body Fat Prediction Dataset	Kaggle	15	251
Heart Disease Dataset Comprehensive	Ieee-dataport.org	11	1190

3.2. Steps for Generating a Hybrid Model

The research process involves the following steps as shown in Figure 1:

Figure 1. Research Process.

3.2.1. Data Partitioning (Separating Dependent and Independent Attributes)

Partitioning the available datasets into two parts, with one part containing the dependent attributes and the other part containing the independent attributes, as illustrated in Figure 2.

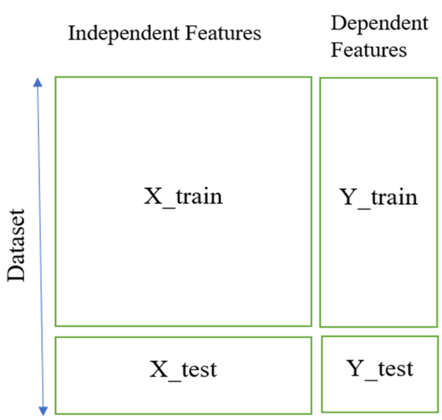

Figure 2. Partitioning of Dataset.

3.2.2. Dataset Split

The conventional and straightforward strategy for dividing the modeling dataset, as suggested in reference [47], involves putting the training set in the ratio of two-thirds of the data points, while the remaining one-third is allocated to the testing set. We divided the dataset into two sets, with 70% of the data assigned to the training set and the remaining 30% to the testing set.

3.2.3 Model Generation

Utilizing both the dependent and independent attributes from the training set to generate a model.

3.2.4. Predictive Analysis

Employing different classifiers to predict the values of the dependent attributes in the testing set using the independent attributes as input.

3.2.5. Performance Evaluation

Assess the model's performance by comparing the predicted values with the available actual values.

3.2.6. Model Effectiveness Assessment

Assessing the effectiveness of these classifiers to determine the optimal model for predicting chronic diseases.

3.3. Foundation of Proposed Framework

The proposed model relies on key foundational techniques, namely linguistic fuzzification processes, principal component analysis, and SVM (as depicted in Figure 3). This section elaborates on the steps of the proposed model.

PHASE-I: DATA PREPARATION:

1. Dataset Selection and Missing Value Imputation:

The study utilizes five datasets namely the Pima Indian Diabetes Dataset, Heart Stroke prediction dataset, Heart Failure Prediction Dataset, Body Fat Prediction Dataset, and Heart Disease Dataset comprehensively outlined in Table 3. The Source of the datasets, the number of Attributes, and the records in each dataset are elaborated in Table 3. These datasets undergo initial cleaning and preprocessing procedures to eliminate missing and null values. Then these datasets were used for carrying out the data pre-processing steps.

2. Data Normalization using Min-Max Tech:

The Min-Max technique is used to scale the features within a predetermined range, preserving the relative relationships between variables. It also ensures uniformity and consistency in data distribution across different features, facilitating optimal model performance.

3. Train-Test Set Preparation:

The available datasets are partitioned into two parts, with one part containing the dependent attributes and the other part containing the independent attributes, as illustrated in Figure 2. The conventional and straightforward strategy for dividing the modeling dataset, as suggested in [47], involves putting the training set in the ratio of two-thirds of the data points, while the remaining one-third is allocated to the testing set. We divided the dataset into two sets, with 70% of the data assigned to the training set and the remaining 30% to the testing set.

PHASE-II: FUZZIFICATION PROCESS:

As dataset is complex and contains a large number of numeric values in some attributes such as heart rate, blood pressure, etc. To make it simple, fuzzification is added at the preprocessing step to make the numeric dataset into crisp form. It helps in making the input simple for training the model. During the fuzzification process, each observation is assigned a membership degree in a fuzzy set. This process involves assigning an observation a degree of membership in one or more fuzzy sets, which transforms numerical data sets into linguistic representations. The triangular membership function is used based on the characteristics of the input data and the application requirements. This approach facilitates the development of an intuitive understanding of the underlying patterns in the data.

Following the feature preprocessing in the data preparation step, the datasets undergo a fuzzification process. This process involves transforming each feature of an input pattern into its corresponding linguistic membership value.

Fuzzification assigns an affiliation level to various fuzzy sets for each observation. This approach allows us to generate a language-based summary of the numerical data and gain insights into the underlying patterns. To address uncertainty and accommodate new input features, linguistic fuzzy expansion expands the number of characteristics into linguistic membership values. Equation (1) represents the pattern of the kth feature in the dataset D.

$$P_m = [F_{m,1}, F_{m,2}, \dots, F_{m,n}], \tag{1}$$

The membership value of the jth feature of the mth pattern denoted as F_m, is determined using Equation (2) Through the utilization of membership functions, the linguistic qualities of the input features are extracted, effectively fuzzifying the original characteristics into appropriate values.

PHASE-III: FEATURE REDUCTION USING PCA:

However, the process of linguistic fuzzification leads to an increase in number of features. Equation (3) demonstrates that the number of features triples after linguistic fuzzification. To address the issue of increased number of features due to linguistic fuzzification, it is essential to extract the characteristics that have a strong influence on the model. Principal Component Analysis (PCA) can be employed to achieve this. By applying PCA, significant features are extracted from the model, while irrelevant and redundant features are discarded. This process ensures that only the characteristics that strongly impact the model are retained.

PHASE-IV: SUPPORT VECTOR MACHINE (SVM) INTEGRATION:

When compared to other ML classifiers, the SVM is the most accurate ML classifier for early detection of heart disease, as per the observations [62,63]. SVM is regarded as a selective classifier, as it is trained by determining optimal hyperplanes that partition the data. It has several benefits, including the capacity to handle high-dimensional data and the usefulness of modeling non-linear decision boundaries in a variety of applications.

When provided with labelled training data, SVM finds the hyperplane that correctly classifies new instances.

$$F_{m,j} = [\mu_{\text{low}}(f_{m,j}), \mu_{\text{medium}}(f_{m,j}), \mu_{\text{high}}(f_{m,j})] \tag{2}$$

$$\begin{aligned} P_i = [&\mu_{\text{low}}(f_{1,i}), \mu_{\text{medium}}(f_{1,i}), \mu_{high}(f_{1,i}), \mu_{low}(f_{2,i}), \mu_{\text{medium}}(f_{2,i}) \\ &\mu_{\text{high}}(f_{2,i}), \dots, \mu_{low}(f_{n,i}), \mu_{\text{medium}}(f_{n,i}), \mu_{high}(f_{n,i})] \end{aligned} \tag{3}$$

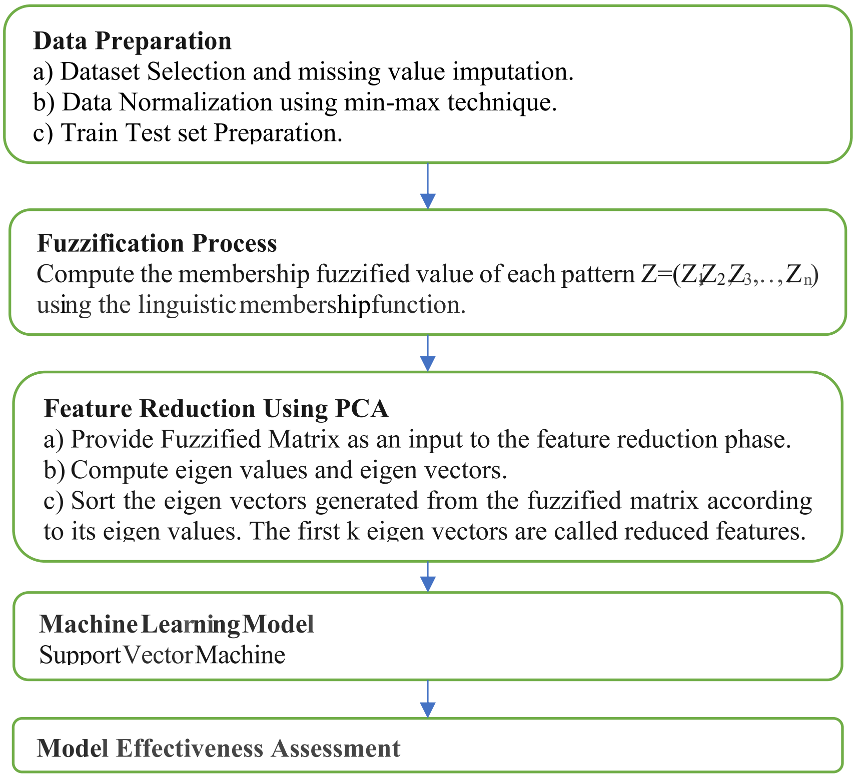

Figure 3. Proposed Framework.

In a two-dimensional space, this hyperplane divides the plane into two regions, each corresponding to a class. SVMs employ support vectors, which are data points closer to the hyperplane, and these support vectors affect the position and orientation of the hyperplane. Maximizing the margin of the classifier is achieved by leveraging these support vectors, and removing them alters the hyperplane's position. Support vectors play a critical role in building SVMs, as they assist in locating an N-dimensional hyperplane that effectively splits the data points. Finding a hyperplane that successfully divides the data points is the aim of the SVM algorithm.

Several hyperplanes can be used to split the data points into different classes, but the goal is to identify the hyperplane that maximizes the margin. By increasing the margin, future data points can be classified with greater confidence. The SVM method increases the separation between the data points and the hyperplane. SVMs are encouraged to locate hyperplanes with a wide margin by hinge loss. The hinged loss function is resistant to data noise, as illustrated in Equation (4), and is utilized to optimize the margins and improve classification accuracy.

$$H(X, Y, f(X)) = \begin{cases} 0, & \text{if } Y * f(X) \geq 1 \\ 1 - Y * f(X), & \text{else} \end{cases} \quad (4)$$

There is no cost when there is a sign of equality between the projected and actual values. However, if the signs differ, a loss value is computed. An easy and effective loss function to optimize is hinge loss. To mitigate overfitting and enhance generalization Train-Test Split & Hyperparameter tuning is done. We started with a coarse grid search or randomized search to narrow down the search space and then used Bayesian Optimization for fine-tuning with parameter range C = 1, Kernal = rbf, and gamma=scale.

SVM is integrated seamlessly with the FL component to harness the complementary strengths of both methodologies for enhanced decision-making and predictive modelling in healthcare settings.

PHASE-V: MODEL VALIDATION AND DEPLOYMENT:

The model's performance is assessed by comparing the predicted values with the available actual values in terms of Precision, Recall, Accuracy, and F1-score.

4. Results

This research focuses on the experimental evaluation of a hybrid model that integrates ML techniques with FL for healthcare classification. The study utilizes five healthcare datasets obtained from various online repositories, covering data related to diabetes, heart stroke, heart failure, and body fat predictions. The empirical findings and evaluations are conducted using the Python 3.8.3 environment with Jupyter Notebook. Seven existing ML algorithms, namely NB, KNN, SVM, DT, RF, Logistic Regression, and PCA-SVM are applied to these datasets to assess various performance parameters. Subsequently, the proposed hybrid Fuzzy-PCA-SVM model is employed on all datasets. The proposed model works well with unseen data.

To evaluate the effectiveness of this model, various performance standards such as accuracy, precision, F1-score, and recall have been considered. In this context, accuracy denotes the proportion of correctly predicted instances among all the available examples. Precision is defined as the proportion of accurate predictions in the occurrences that fall into the positive category. The recall is characterized as the ratio of correctly identified Positive samples to all Positive samples. Using their harmonic mean, the F1 score combines recall and precision. Figure 4 illustrates the comparative analysis of accuracy, while Figure 5, Figure 6, and Figure 7 depict the analysis of precision, recall, and F1-score, respectively. Each classifier's performance is assessed on an individual basis, and all outcomes are dutifully documented for further analysis. Table 4 depicts the results of accuracy, Table 5 depicts the results of precision, Table 6 depicts the results of recall and Table 7 depicts the results of F1-score all five datasets respectively. From the experimental data provided, several key findings and interpretations can be derived:

1. Overall Performance Comparison: Fuzzy-PCA-SVM consistently outperforms other models in terms of accuracy across all datasets. It achieves the highest accuracy scores in four out of the five datasets, indicating its effectiveness in making overall correct predictions.
2. Accuracy Improvement: Fuzzy-PCA-SVM shows an improvement in accuracy compared to other models, especially in datasets such as the Pima Indian Diabetes Dataset (set 1) and Heart Stroke Prediction Dataset (set 2), where it achieves notably higher accuracy scores.
3. Consistent High Performance: Fuzzy-PCA-SVM demonstrates robust performance across multiple datasets, suggesting its reliability in different scenarios. It consistently achieves accuracy scores above 0.85 in most datasets, indicating its effectiveness in handling various classification tasks.
4. Competitive Performance: While other models like Random Forest (RF) and PCA-SVM also show competitive accuracy scores, Fuzzy-PCA-SVM consistently outperforms them, especially in datasets with complex patterns or imbalanced classes.

5. Potential for Further Improvement: The proposed Fuzzy-PCA-SVM model shows promising results, especially in the Heart Disease Dataset Comprehensive (set 5), where it achieves the highest accuracy score of 0.92. This suggests that the model has the potential for further optimization and improvement.

Table 4. Accuracy of five healthcare datasets.

Datasets/Models	**Accuracy**							
	Naïve Bayes	**KNN**	**Logistic Regression**	**SVM**	**DT**	**RF**	**PCA-SVM**	**Fuzzy- PCA-SVM**
Pima Indian Diabetes Dataset (set 1)	0.75	0.74	0.71	0.77	0.73	0.75	0.78	0.79
Heart Stroke Prediction Dataset (set 2)	0.78	0.82	0.81	0.84	0.82	0.84	0.85	0.87
Heart Failure Prediction Dataset (set 3)	0.74	0.77	0.78	0.83	0.82	0.83	0.84	0.85
Body Fat Prediction Dataset (set 4)	0.73	0.79	0.76	0.81	0.8	0.81	0.82	0.83
Heart Disease Dataset Comprehensive (set 5)	0.81	0.84	0.88	0.86	0.83	0.9	0.9	0.92

In summary, the experimental data indicates that Fuzzy-PCA-SVM is the most effective model among the compared algorithms in terms of accuracy, demonstrating consistent high performance across multiple datasets. Its competitive performance and potential for further improvement make it a promising candidate for classification tasks in various domains.

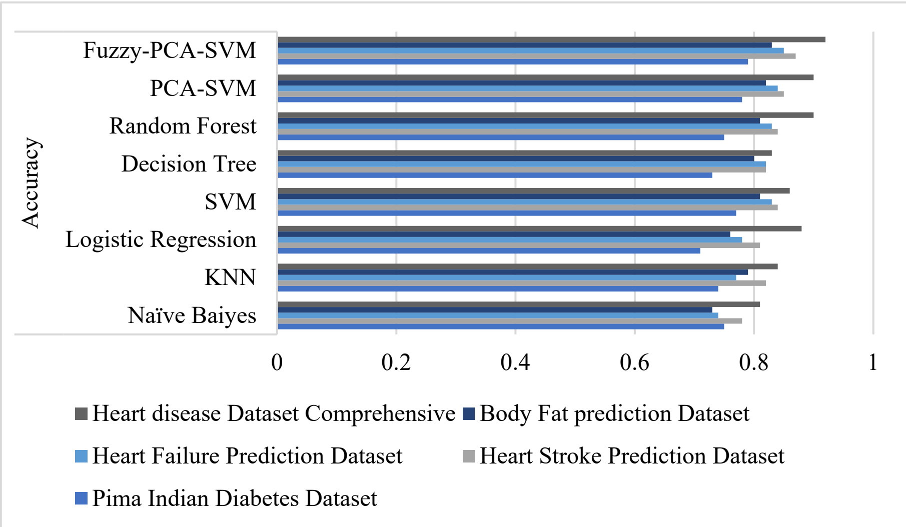

Figure 4. Classifiers Accuracy on five healthcare datasets.

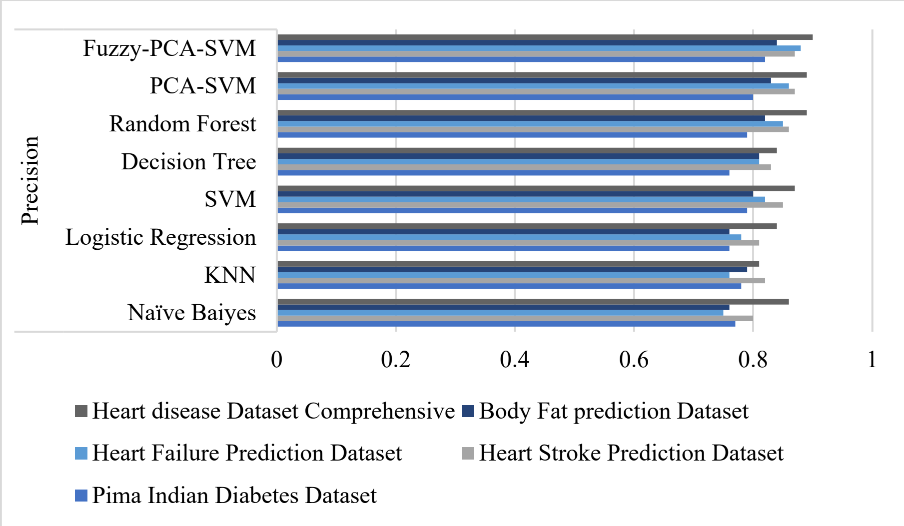

Figure 5. Classifiers Precision on five healthcare datasets.

Table 5. Precision of five healthcare datasets.

Datasets/Models	**Precision**							
	Naïve Bayes	**KNN**	**Logistic Regression**	**SVM**	**DT**	**RF**	**PCA-SVM**	**Fuzzy- PCA-SVM**
Pima Indian Diabetes Dataset (set 1)	0.77	0.78	0.76	0.79	0.76	0.79	0.8	0.82
Heart Stroke Prediction Dataset (set 2)	0.8	0.82	0.81	0.85	0.83	0.86	0.87	0.87
Heart Failure Prediction Dataset (set 3)	0.75	0.76	0.78	0.82	0.81	0.85	0.86	0.88
Body Fat Prediction Dataset (set 4)	0.76	0.79	0.76	0.8	0.81	0.82	0.83	0.84
Heart Disease Dataset Comprehensive (set 5)	0.86	0.81	0.84	0.87	0.84	0.89	0.89	0.9

From Table 5, following inferences can be drawn regarding the precision performance of different models on various datasets:

1. Fuzzy-PCA-SVM consistently achieves the highest precision scores across all datasets, indicating its effectiveness in making accurate positive predictions while minimizing false positives.
2. SVM and PCA-SVM generally exhibit high precision scores across datasets, suggesting they are robust in making accurate positive predictions.
3. Naïve Bayes and DT tend to achieve lower precision scores compared to other models, indicating they may be prone to a higher rate of false positive predictions.
4. The Heart Disease Dataset Comprehensive (set 5) consistently yields higher precision scores across all models, suggesting it may have a more balanced distribution of positive and negative samples, making it easier for models to make accurate positive predictions.

Overall, the impact of the models on the datasets varies, but SVM, Random Forest (RF), PCA-SVM, and Fuzzy-PCA-SVM consistently demonstrate strong performance across different datasets, indicating their effectiveness in handling various prediction tasks. These models tend to capture the complexities of the datasets well and provide high precision in their predictions. Naïve Bayes and Logistic Regression, while competitive, generally exhibit slightly lower precision scores compared to the more sophisticated models across most datasets.

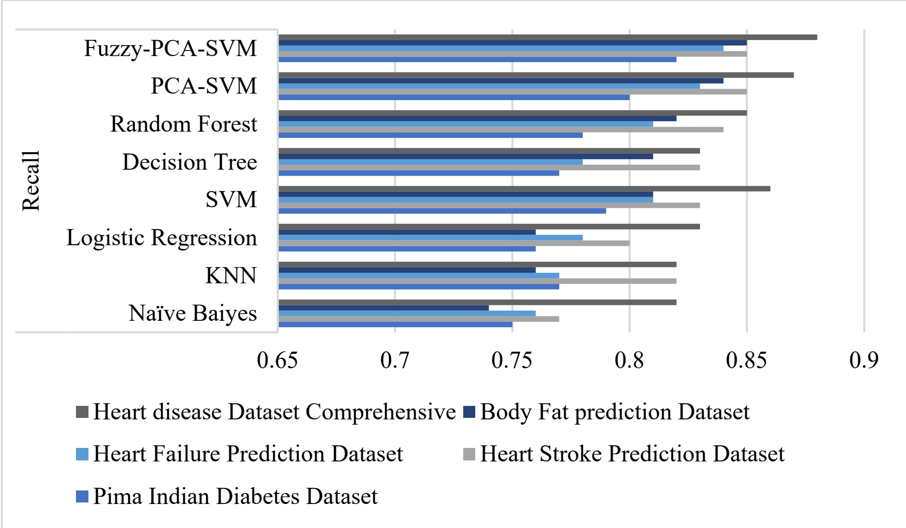

Figure 6. Classifiers Recall on five healthcare datasets.

Table 6. Recall of five healthcare datasets.

Datasets/Models	Recall							
	Naïve Bayes	**KNN**	**Logistic Regression**	**SVM**	**DT**	**RF**	**PCA-SVM**	**Fuzzy- PCA-SVM**
Pima Indian Diabetes Dataset (set 1)	0.75	0.77	0.76	0.79	0.77	0.78	0.8	0.82
Heart Stroke Prediction Dataset (set 2)	0.77	0.82	0.8	0.83	0.83	0.84	0.85	0.85
Heart Failure Prediction Dataset (set 3)	0.76	0.77	0.78	0.81	0.78	0.81	0.83	0.84
Body Fat Prediction Dataset (set 4)	0.74	0.76	0.76	0.81	0.81	0.82	0.84	0.85
Heart Disease Dataset Comprehensive (set 5)	0.82	0.82	0.83	0.86	0.83	0.85	0.87	0.88

From Table 6, several inferences can be drawn regarding the recall performance of different models on various datasets:

1. Fuzzy-PCA-SVM: Generally leads in recall across most datasets, indicating it effectively identifies relevant data points while minimizing false negatives, and exhibits the effectiveness of the hybrid approach in improving recall performance.
2. .SVM and RF: Consistently rank high in recall, demonstrating their ability to capture complex relationships and avoid missing true positives.
3. PCA-SVM: Performs well in most datasets, suggesting dimensionality reduction is beneficial for recall in some cases.
4. Naïve Bayes and KNN: Show decent recall but are sometimes outperformed by other models, potentially due to limitations in handling complex relationships or specific data characteristics.

Overall, the recall data suggests Fuzzy-PCA-SVM might be a strong choice for tasks prioritizing minimizing false negatives, particularly in complex datasets with potentially fuzzy features. However, a broader analysis considering all evaluation metrics, statistical significance, and detailed model information is crucial for robust decisions.

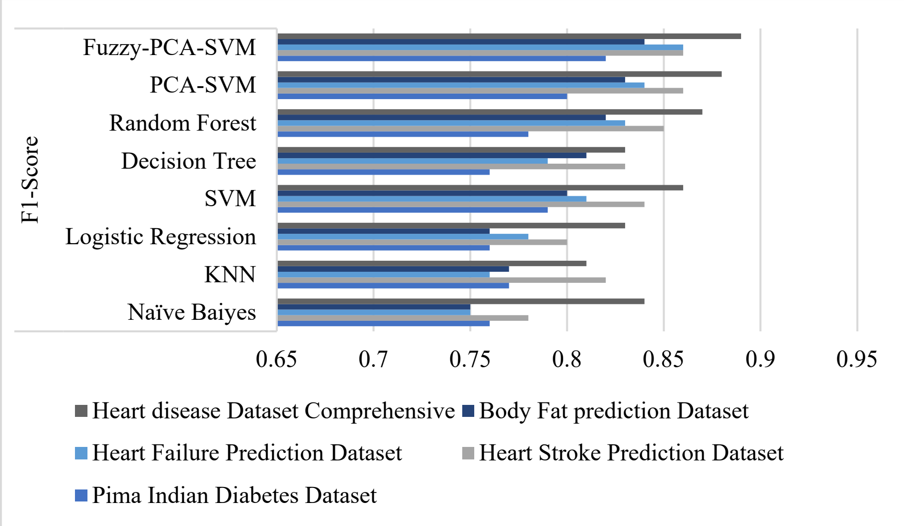

Figure 7. Classifiers F1-Score on five healthcare datasets.

Table 7. F1-Score of five healthcare datasets.

Datasets/Models	F1-Score							
	Naïve Bayes	**KNN**	**Logistic Regression**	**SVM**	**DT**	**RF**	**PCA-SVM**	**Fuzzy- PCA-SVM**
Pima Indian Diabetes Dataset (set 1)	0.76	0.77	0.76	0.79	0.76	0.78	0.8	0.82
Heart Stroke Prediction Dataset (set 1)	0.78	0.82	0.8	0.84	0.83	0.85	0.86	0.86
Heart Failure Prediction Dataset (set 3)	0.75	0.76	0.78	0.81	0.79	0.83	0.84	0.86
Body Fat Prediction Dataset (set 4)	0.75	0.76	0.78	0.81	0.79	0.83	0.84	0.86
Heart Disease Dataset Comprehensive (set 5)	0.84	0.81	0.83	0.86	0.83	0.87	0.88	0.89

From the above Table 7, the following inferences can be drawn regarding the performance of different models on various datasets:

1. Fuzzy-PCA-SVM: Achieves the highest overall F1-score across most datasets, suggesting a good balance between precision and recall, minimizing both false positives and negatives.
2. SVM and RF: Consistently rank high in F1-score, demonstrating their ability to achieve both high precision and recall in diverse classification tasks.
3. PCA-SVM: Performs well in most datasets, indicating dimensionality reduction is beneficial for balanced performance in some cases.
4. Naïve Bayes and KNN: Show decent F1-score but are sometimes outperformed by other models, potentially due to limitations in specific datasets or complex relationships.
5. The Heart Disease Dataset Comprehensive (set 5) consistently yields higher F1-scores across all models, suggesting it may be a more discriminative dataset for classification tasks compared to other datasets.

Overall, the F1-score data suggests Fuzzy-PCA-SVM might be a strong choice for tasks requiring a balance between precision and recall, particularly in complex datasets with potentially fuzzy features.

5. Discussion

Following are some final interpretations, inferences, and key findings regarding the performance of various models on different datasets:

- **Overall Performance:**
  1. Fuzzy-PCA-SVM: Consistently ranks high or achieves the best score across all metrics and most datasets, suggesting it's a strong general-purpose model for diverse classification tasks.
  2. SVM and RF: Often perform competitively across most metrics and datasets, indicating their

versatility and effectiveness in handling various data characteristics.

3. PCA-SVM: Shows potential benefits in some datasets, suggesting dimensionality reduction can be helpful for specific data, but its performance varies.
4. Naïve Bayes and KNN: Generally, show lower performance compared to other models across most metrics and datasets, suggesting limitations in handling complex relationships or specific data types.

- **Dataset-Specific Observations:**
  1. Pima Indian Diabetes: Fuzzy-PCA-SVM leads in most metrics, followed by SVM and RF.
  2. Heart Stroke Prediction: Fuzzy-PCA-SVM and SVM are neck-and-neck, with RF close behind in most metrics.
  3. Heart Failure Prediction: Fuzzy-PCA-SVM leads in most metrics, followed by SVM and RF.
  4. Body Fat Prediction: Fuzzy-PCA-SVM again excels in most metrics, followed by SVM and RF.
  5. Heart Disease Dataset Comprehensive: Fuzzy-PCA-SVM demonstrates clear dominance in most metrics, followed by SVM and RF.
- **Key Inferences:**
  1. Handling Fuzzy Features: If datasets involve inherent fuzziness, Fuzzy-PCA-SVM might be particularly effective in incorporating such characteristics for better overall performance.
  2. Complex Data: Fuzzy-PCA-SVM's consistent leadership suggests it effectively handles complex data with both numerical and potentially fuzzy features.
  3. Dimensionality Reduction: While PCA-SVM shows benefits in some cases, its impact varies depending on the dataset, suggesting careful evaluation for specific scenarios.
  4. Non-linear Relationships: SVM and RF's strong performance in many datasets indicates their ability to capture complex interactions, potentially beneficial for non-linear relationships.
- **Limitations and Further Analysis:**
  1. **Missing Information:** Detailed information about datasets, features, and model configurations is crucial for drawing definitive conclusions. It is highly desirable to advance this study further, emphasizing real-world datasets rather than solely relying on theoretical frameworks.
  2. **Statistical Significance:** Performing statistical tests like paired t-tests or ANOVA can solidify observed performance differences between models.
  3. **Visualization:** Techniques like confusion matrices or ROC curves could provide deeper insights into model behavior for specific classes or prediction probabilities.

Overall, this combined analysis suggests Fuzzy-PCA-SVM emerges as a promising model for diverse classification tasks, particularly when dealing with complex or potentially fuzzy data. However, considering all evaluation metrics, statistical significance, and detailed model information is crucial for robust decisions.

A complete list of abbreviations is shown in Appendix *I.*

6. Conclusions and Future Work

In conclusion, this research paper presents a comprehensive investigation into the application of ML techniques for disease diagnosis, emphasizing the necessity of integrating both linguistic and numeric inputs to enhance diagnostic accuracy and robustness. Through an empirical evaluation using diverse healthcare datasets, the study introduces a hybrid ML-fuzzy logic (FL) model, Fuzzy-PCA-SVM, and compares its performance with seven existing ML algorithms. Five healthcare datasets related to diabetes, heart stroke, heart failure, and body fat predictions were utilized for experimental evaluation. Existing ML algorithms, including DT, RF, KNN, Logistic Regression, NB, SVM and PCA-SVM, were applied to the datasets to assess performance measures such as accuracy, F1-score, recall, and precision. The results demonstrate the superiority of the proposed hybrid model, particularly in handling complex and potentially fuzzy data, as evidenced by its consistently high performance across multiple metrics and datasets. While SVM and RF also exhibit competitive performance, the study highlights the potential of integrating FL systems with ML techniques to achieve superior classification outcomes. Furthermore, the paper identifies limitations and areas for further analysis, emphasizing the importance of robust evaluation methodologies and the need for continued advancements in feature selection techniques and real-world dataset utilization in healthcare analytics research. Overall, the findings underscore the significance of leveraging hybrid ML-FL models like Fuzzy-PCA-SVM for enhanced disease diagnosis and pave the way for future research aimed at improving healthcare classification outcomes.

Author Contributions

K.G. Conceptualization, data curation, formal analysis, investigation, methodology, software, visualization, validation, writing - original draft, writing—review & editing. P.K.: Writing—original draft, conceptualization, data curation, methodology, software, validation, supervision. S.U.: Writing—original draft, conceptualization, data curation, methodology, software, validation, supervision. M.P.: conceptualization and supervision. S.A.: conceptualization and supervision. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Data Availability Statement

Datasets are publically available on Kaggle and Ieee-dataport.org websites.

References

1. G. Uysal and M. Ozturk, "Hippocampal atrophy based Alzheimer's disease diagnosis via machine learning methods," *J. Neurosci. Methods*, vol. 337, p. 108669, 2020.
2. P. Srihari and J. Harikiran, "A Deep Learning Based Hybrid Approach for Human Physical Activity Recognition in Thermal Imaging," *Int. J. Comput. Inf. Syst. Ind. Manag. Appl.*, vol. 14, pp. 9–9, 2022.
3. A. Salih and A. Abraham, "A review of ambient intelligence assisted healthcare monitoring," *Int. J. Comput. Inf. Syst. Ind. Manag. IJCISIM*, vol. 5, pp. 741–750, 2013.
4. K. S. Kaswan and J. S. Dhatterwal, "The Use of Machine Learning for Sustainable and Resilient Buildings," in *Digital Cities Roadmap*, 1st ed., A. Solanki, A. Kumar, and A. Nayyar, Eds., Wiley, 2021, pp. 1–62. doi: 10.1002/9781119792079.ch1.
5. F. C. Morabito *et al.*, "Deep convolutional neural networks for classification of mild cognitive impaired and Alzheimer's disease patients from scalp EEG recordings," in *2016 IEEE 2nd International Forum on Research and Technologies for Society and Industry Leveraging a better tomorrow (RTSI)*, IEEE, 2016, pp. 1–6. [Online]. Available at: https://ieeexplore.ieee.org/abstract/document/7740576/ (Accessed on date: 13 February 2024).
6. M. D. Bloice and A. Holzinger, "A Tutorial on Machine Learning and Data Science Tools with Python," in *Machine Learning for Health Informatics*, vol. 9605, A. Holzinger, Ed., in Lecture Notes in Computer Science, vol. 9605. , Cham: Springer International Publishing, 2016, pp. 435–480. doi: 10.1007/978-3-319-50478-0_22.
7. I. Ullah, H. Y. Youn, and Y.-H. Han, "Integration of type-2 fuzzy logic and Dempster–Shafer Theory for accurate inference of IoT-based health-care system," *Future Gener. Comput. Syst.*, vol. 124, pp. 369–380, 2021.
8. S. S. Tomar and V. G. Narendra, "Python-Based Fuzzy Classifier for Cashew Kernels," in *Soft Computing for Problem Solving*, vol. 816, J. C. Bansal, K. N. Das, A. Nagar, K. Deep, and A. K. Ojha, Eds., in Advances in Intelligent Systems and Computing, vol. 816. , Singapore: Springer Singapore, 2019, pp. 365–374. doi: 10.1007/978-981-13-1592-3_28.
9. K. Gupta, N. Hooda, P. Mittal, S. U. Bhasin, and R. kumar, "Schematic review of sentiment analysis techniques," presented at the 23rd Intelligent Systems Design and Applications, Olten, Switzerland; Porto, Portugal; Vilnius, Lithuania and Kochi, India: Springer.
10. V. Sharma, A. Bajaj, and A. Abraham, "Machine Learning Techniques for Electronic Health Records: Review of a Decade of Research. | International Journal of Computer Information Systems & Industrial Management Applications | EBSCOhost." [Online]. Available at: https://openurl.ebsco.com/contentitem/gcd:174486614?sid=ebsco:plink:crawler&id=ebsco:gcd:174486614 (Accessed on date: 28 March 2024).
11. I. Boudali, "Data-Driven Prediction of Complications Risks in Cancer Patients: Machine Learning based approach. | International Journal of Computer Information Systems & Industrial Management Applications | EBSCOhost." Available at: https://openurl.ebsco.com/contentitem/gcd:174486617?sid=ebsco:plink:crawler&id=ebsco:gcd:174486617 (Accessed on date: 28 March 2024).
12. M. AlJame, I. Ahmad, A. Imtiaz, and A. Mohammed, "Ensemble learning model for diagnosing COVID-19 from routine blood tests," *Inform. Med. Unlocked*, vol. 21, p. 100449, 2020.
13. Y. A. Nanehkaran *et al.*, "Anomaly detection in heart disease using a density-based unsupervised approach," *Wirel. Commun. Mob. Comput.*, vol. 2022, 2022. [Online]. Available at: https://www.hindawi.com/journals/wcmc/2022/6913043/ (Accessed on date: 2 February 2024).
14. M. Liu, X. Shen, and W. Pan, "Deep reinforcement learning for personalized treatment recommendation," *Stat. Med.*, vol. 41, no. 20, pp. 4034–4056, Sep. 2022, doi: 10.1002/sim.9491.
15. Z. Zhang, "Reinforcement learning in clinical medicine: a method to optimize dynamic treatment regime over time," *Ann. Transl. Med.*, vol. 7, no. 14, 2019, [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6694251/ (Accessed on date: 2 February 2024).
16. G. Litjens et al., "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, 2017.

17. M. Tsuneki, “Deep learning models in medical image analysis,” *J. Oral Biosci.*, vol. 64, no. 3, pp. 312–320, 2022.
18. J. Schuetter, S. Mishra, L. Lin, and D. Chandramohan, “Ensemble Learning: A Robust Paradigm for Data-Driven Modeling in Unconventional Reservoirs,” In *Proceedings of the 7th Unconventional Resources Technology Conference*, Denver, Colorado, USA: American Association of Petroleum Geologists, 2019. doi: 10.15530/urtec-2019-929.
19. P. Gupta, P. Malhotra, J. Narwariya, L. Vig, and G. Shroff, “Transfer Learning for Clinical Time Series Analysis Using Deep Neural Networks,” *J. Healthc. Inform. Res.*, vol. 4, no. 2, pp. 112–137, Jun. 2020, doi: 10.1007/s41666-019-00062-3.
20. M. Subramanian, V. E. Sathishkumar, J. Cho, and K. Shanmugavadivel, “Learning without forgetting by leveraging transfer learning for detecting COVID-19 infection from CT images,” *Sci. Rep.*, vol. 13, no. 1, p. 8516, 2023.
21. V. Olsson, W. Tranheden, J. Pinto, and L. Svensson, “Classmix: Segmentation-based data augmentation for semi-supervised learning,” in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021, pp. 1369–1378. [Online]. Available at: http://openaccess.thecvf.com/content/WACV2021/html/Olsson_ClassMix_Segmentation-Based_Data_Augmentation_for_Semi-Supervised_Learning_WACV_2021_paper.html (Accessed on date: 2 February 2024).
22. I. Triguero, S. García, and F. Herrera, “Self-labeled techniques for semi-supervised learning: taxonomy, software and empirical study,” *Knowl. Inf. Syst.*, vol. 42, no. 2, pp. 245–284, Feb. 2015, doi: 10.1007/s10115-013-0706-y.
23. A. Rafiei, R. Moore, S. Jahromi, F. Hajati, and R. Kamaleswaran, “Meta-learning in healthcare: A survey.” arXiv, Aug. 05, 2023. [Online]. Available at: http://arxiv.org/abs/2308.02877 (Accessed on date: 2 February 2024).
24. K. Singh and D. Malhotra, “Meta-Health: Learning-to-Learn (Meta-learning) as a Next Generation of Deep Learning Exploring Healthcare Challenges and Solutions for Rare Disorders: A Systematic Analysis,” *Arch. Comput. Methods Eng.*, vol. 30, no. 7, pp. 4081–4112, Sep. 2023, doi: 10.1007/s11831-023-09927-8.
25. Z. C. Antoniou, A. S. Panayides, M. Pantzaris, A. G. Constantinides, C. S. Pattichis, and M. S. Pattichis, “Real-time adaptation to time-varying constraints for medical video communications,” *IEEE J. Biomed. Health Inform.*, vol. 22, no. 4, pp. 1177–1188, 2017.
26. U. Pawar, D. O’Shea, S. Rea, and R. O’Reilly, “Incorporating Explainable Artificial Intelligence (XAI) to aid the Understanding of Machine Learning in the Healthcare Domain.,” in *AICS*, 2020, pp. 169–180. [Online]. Available at: https://www.researchgate.net/profile/Ruairi-Oreilly/publication/346717871_Incorporating_Explainable_Artificial_Intelligence_XAI_to_aid_the_Understanding_of_Machine_Learning_in_the_Healthcare_Domain/links/5fd9374e299bf1408811d849/Incorporating-Explainable-Artificial-Intelligence-XAI-to-aid-the-Understanding-of-Machine-Learning-in-the-Healthcare-Domain.pdf (Accessed on date: 2 February 2024)..
27. M. Chevallier, N. Rogovschi, F. Boufarès, N. Grozavu, and C. Clairmont, “Detecting near duplicate dataset with machine learning,” *Int. J. Comput. Inf. Syst. Ind. Manag. Appl.*, 2022. [Online]. Available at: https://hal.science/hal-03722301/ (Accessed on date: 28 March 2024).
28. H. O. Alanazi, A. H. Abdullah, and K. N. Qureshi, “A critical review for developing accurate and dynamic predictive models using machine learning methods in medicine and health care,” *J. Med. Syst.*, vol. 41, pp. 1–10, 2017.
29. N. Kushwaha and M. Pant, “Fuzzy magnetic optimization clustering algorithm with its application to health care,” *J. Ambient Intell. Humaniz. Comput.*, pp. 1–10, 2018.
30. S. Uddin, A. Khan, M. E. Hossain, and M. A. Moni, “Comparing different supervised machine learning algorithms for disease prediction,” *BMC Med. Inform. Decis. Mak.*, vol. 19, no. 1, p. 281, Dec. 2019, doi: 10.1186/s12911-019-1004-8.
31. S. Graham *et al.*, “Artificial intelligence for mental health and mental illnesses: an overview,” *Curr. Psychiatry Rep.*, vol. 21, pp. 1–18, 2019.
32. S. Shukla, M. F. Hassan, L. T. Jung, A. Awang, and M. K. Khan, “A 3-Tier Architecture for Network Latency Reduction in Healthcare Internet-of-Things Using Fog Computing and Machine Learning,” In *Proceedings of the 2019 8th International Conference on Software and Computer Applications*, Penang Malaysia: ACM, Feb. 2019, pp. 522–528. doi: 10.1145/3316615.3318222.
33. M. Bhatt and T. Kasbe, “A survey on chronic kidney disease diagnosis using fuzzy logic,” in *2019 IEEE International Symposium on Smart Electronic Systems (iSES)(Formerly iNiS)*, IEEE, 2019, pp. 252–256. [Online]. Available at: https://ieeexplore.ieee.org/abstract/document/9002448/ (Accessed on date: 13 February 2024).
34. F. Z. Benchara and M. Youssfi, “A new distributed type-2 fuzzy logic method for efficient data science models of medical informatics,” *Adv. Fuzzy Syst.*, vol. 2020, pp. 1–10, 2020.
35. G. Casalino, R. Grassi, M. Iannotta, V. Pasquadibisceglie, and G. Zaza, “A hierarchical fuzzy system for risk assessment of cardiovascular disease,” In *2020 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS)*, IEEE, 2020, pp. 1–7. [Online]. Available at: https://ieeexplore.ieee.org/ abstract/document/9122750/ (Accessed on date: 13 February 2024).
36. N. A. Omoregbe, I. O. Ndaman, S. Misra, O. O. Abayomi-Alli, R. Damaševičius, and A. Dogra, “Text messaging-based medical diagnosis using natural language processing and fuzzy logic,” *J. Healthc. Eng.*, vol. 2020, pp. 1–14, 2020.
37. A. H. Zubar and R. Balamurugan, “Green Computing Process and its Optimization Using Machine Learning Algorithm in Healthcare Sector,” *Mob. Netw. Appl.*, vol. 25, no. 4, pp. 1307–1318, Aug. 2020, doi: 10.1007/s11036-020-01549-9.

38. G. T. Reddy, M. P. K. Reddy, K. Lakshmanna, D. S. Rajput, R. Kaluri, and G. Srivastava, "Hybrid genetic algorithm and a fuzzy logic classifier for heart disease diagnosis," *Evol. Intell.*, vol. 13, no. 2, pp. 185–196, Jun. 2020, doi: 10.1007/s12065-019-00327-1.
39. M. Jayalakshmi *et al.*, "Fuzzy logic-based health monitoring system for COVID'19 patients," *Comput. Mater. Contin.*, vol. 67, no. 2, pp. 2431–2447, 2021.
40. H. Harb, A. Mansour, A. Nasser, E. M. Cruz, and I. de la Torre Diez, "A sensor-based data analytics for patient monitoring in connected healthcare applications," *IEEE Sens. J.*, vol. 21, no. 2, pp. 974–984, 2020.
41. A. Duggento, A. Conti, M. Guerrisi, and N. Toschi, "Detection of abnormal phonocardiograms through the Mel-frequency ceptrum and convolutional neural networks," in *2020 11th Conference of the European Study Group on Cardiovascular Oscillations (ESGCO)*, IEEE, 2020, pp. 1–2. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9158167/ (Accessed on date: 14 February 2024)..
42. B. Mostafa, A. Miry, and T. Salman, "Healthcare Monitoring and Analytic System Based Internet of Thing," *Iraqi J Electr Electron Eng*, pp. 30–36, 2020.
43. M. Annamalal, X. M. Jesintha, and M. Nithya, "Smart IOT based healthcare monitoring and decision-making system using augmented data recognition algorithm," *Turk. J. Comput. Math. Educ.*, vol. 12, no. 11, pp. 1971–1979, 2021.
44. K. Hameed, I. S. Bajwa, S. Ramzan, W. Anwar, and A. Khan, "An intelligent IoT based healthcare system using fuzzy neural networks," *Sci. Program.*, vol. 2020, pp. 1–15, 2020.
45. M. Alshamrani, "Remote Intelligent Healthcare System Based on the ROCKET Technique," *Arab. J. Sci. Eng.*, vol. 46, no. 9, pp. 9263–9277, Sep. 2021, doi: 10.1007/s13369-021-05805-2.
46. A. Rahman, M. Rahman, D. Kundu, M. R. Karim, S. S. Band, and M. Sookhak, "Study on IoT for SARS-CoV-2 with healthcare: present and future perspective," *Math. Biosci. Eng.*, vol. 18, no. 6, pp. 9697–9726, 2021.
47. A. Sharma, "The Role of IoT in the Fight Against Covid-19 to Restructure the Economy," in *HCI International 2021 - Late Breaking Papers: HCI Applications in Health, Transport, and Industry*, vol. 13097, C. Stephanidis, V. G. Duffy, H. Krömker, F. Fui-Hoon Nah, K. Siau, G. Salvendy, and J. Wei, Eds., in Lecture Notes in Computer Science, vol. 13097., Cham: Springer International Publishing, 2021, pp. 140–156. doi: 10.1007/978-3-030-90966-6_11.
48. N. P. Tigga and S. Garg, "Prediction of type 2 diabetes using machine learning classification methods," *Procedia Comput. Sci.*, vol. 167, pp. 706–716, 2020.
49. A. Thilagavathy, S. Meenakshi, V. Vijayabhaskar, M. D. Babu, S. Kumari, and M. A. Gunavathie, "An Efficient Health Monitoring Method Using Fuzzy Inference System via Cloud," *Indian J Sci Technol*, vol. 14, pp. 2145–2151, 2021.
50. S. Vyas, S. Gupta, D. Bhargava, and R. Boddu, "Fuzzy Logic System Implementation on the Performance Parameters of Health Data Management Frameworks," *J. Healthc. Eng.*, vol. 2022, 2022, [Online]. Available: https://www.hindawi.com/journals/jhe/2022/9382322/ (Accessed on date: 14 February 2024).
51. K. M. Zobair, L. Sanzogni, L. Houghton, and M. Z. Islam, "Forecasting care seekers satisfaction with telemedicine using machine learning and structural equation modeling," *PloS One*, vol. 16, no. 9, p. e0257300, 2021.
52. O. H. Salman, Z. Taha, M. Q. Alsabah, Y. S. Hussein, A. S. Mohammed, and M. Aal-Nouman, "A review on utilizing machine learning technology in the fields of electronic emergency triage and patient priority systems in telemedicine: Coherent taxonomy, motivations, open research challenges and recommendations for intelligent future work," *Comput. Methods Programs Biomed.*, vol. 209, p. 106357, 2021.
53. T. Malapane, W. Doorsamy, and B. S. Paul, "An intelligent IoT-based health monitoring system," in *2020 International Conference on Intelligent Data Science Technologies and Applications (IDSTA)*, IEEE, 2020, pp. 95–100. Available at: https://ieeexplore.ieee.org/abstract/document/ 9264102/ (Accessed on date: 14 February 2024).
54. J. McNulty, R. B. Reilly, T. E. Taylor, S. M. O'Dwyer, R. W. Costello, and Y. Zigel, "Automatic Audio-Based Classification of Patient Inhaler Use: A Pharmacy Based Study," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, 2019, pp. 2606–2609. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8857132/ (Accessed on date: 14 February 2024).
55. J. N. Saeed and S. Y. Ameen, "Smart Healthcare for ECG Telemonitoring System," *J. Soft Comput. Data Min.*, vol. 2, no. 2, pp. 75–85, 2021.
56. A. Esteva *et al.*, "Deep learning-enabled medical computer vision," *NPJ Digit. Med.*, vol. 4, no. 1, p. 5, 2021.
57. M. Owais, M. Arsalan, J. Choi, T. Mahmood, and K. R. Park, "Artificial intelligence-based classification of multiple gastrointestinal diseases using endoscopy videos for clinical diagnosis," *J. Clin. Med.*, vol. 8, no. 7, p. 986, 2019.
58. A. Nithya, A. Appathurai, N. Venkatadri, D. R. Ramji, and C. A. Palagan, "Kidney disease detection and segmentation using artificial neural network and multi-kernel k-means clustering for ultrasound images," *Measurement*, vol. 149, p. 106952, 2020.
59. K. Gupta, P. Kumar, and S. Upadhyaya, "Predictive Analysis in Healthcare: A Survey," in *2022 Seventh International Conference on Parallel, Distributed and Grid Computing (PDGC)*, IEEE, 2022, pp. 156–161. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10053246/ (Accessed on date: 14 February 2024).
60. S. Balakrishna, M. Thirumaran, and V. K. Solanki, "IoT Sensor Data Integration in Healthcare using Semantics and Machine Learning Approaches," in *A Handbook of Internet of Things in Biomedical and Cyber Physical System*, vol. 165, V. E. Balas, V. K. Solanki, R. Kumar, and Md. A. R. Ahad, Eds., in Intelligent Systems Reference Library, vol. 165. , Cham: Springer International Publishing, 2020, pp. 275–300. doi: 10.1007/978-3-030-23983-1_11.

61. H. F. Atlam, R. J. Walters, G. B. Wills, and J. Daniel, "Fuzzy Logic with Expert Judgment to Implement an Adaptive Risk-Based Access Control Model for IoT," *Mob. Netw. Appl.*, vol. 26, no. 6, pp. 2545–2557, Dec. 2021, doi: 10.1007/s11036-019-01214-w.
62. M. Kumari and S. Godara, "Comparative study of data mining classification methods in cardiovascular disease prediction," *Ijcst*, vol. 2, no. 2, pp. 304–308, 2011.
63. M. A. Jabbar, B. L. Deekshatulu, and P. Chandra, "Computational intelligence technique for early diagnosis of heart disease," in *2015 IEEE International Conference on Engineering and Technology (ICETECH)*, IEEE, 2015, pp. 1–6. [Online]. Available at: https://ieeexplore.ieee.org/abstract/ document/7275001/ (Accessed on date: 14 February 2024).

Author Biographies

Kirti Gupta was born in Karnal district of Haryana. She received her bachelor (B.Tech.) in Computer Engineering from Kurukshetra University, Haryana in 2012. She received her master's degree (M.Tech.) in Computer Science & Engineering from department of Computer Science & Applications, Kurukshetra University in 2014 and now she is pursuing PhD in Computer Science & Applications from the department of Computer Science & Applications, Kurukshetra University, Kurukshetra (Haryana) and her current research work is in Machine Learning and Fuzzy Logic.
Email: kirtigupta811@gmail.com

Pardeep Kumar received his PhD in Computer Science & Applications, M.Sc. (Computer Science) and M.Sc. (Statistics) degree from Kurukshetra University, Kurukshetra. Presently he is working as Associate Professor in Kurukshetra University, Kurukshetra. His research interest lies in Optimization, Cloud Computing, Network Routing and Soft Computing. He has published more than 75 research papers in referred journals and international conferences.
Email: pmittal@kuk.ac.in

Shuchita Upadhyaya has received her PhD in Computer Science and Applications. Currently, she is working as a Professor and has 34 years of teaching experience in the Department of Computer Science and Applications, at Kurukshetra University. Her research interests are Computer Networks & Data Communication, Internet Technologies, Network Security, Wireless and Adhoc networks, Wireless Sensor Networks, Data Mining, Machine Learning, and Fuzzy Logic. Around 120 research papers have been published by her in various journals/conferences.
Email: shuchita_bhasin@yahoo.com

Monika Poriye, is currently working as an Assistant Professor in Department of Computer Science and Applications, Kurukshetra University. She has completed her doctorate in Computer Science & Applications, M.Tech. and M.Sc. from Kurukshetra University, Kurukshetra. Her area of interest is Information Security, Web Development, Cloud Computing, Machine Learning etc. Dr. Monika has published more than 40 research papers in various journals/conferences.
Email: monikaporiye@gmail.com

Shalini Aggarwal is currently working as an Assistant Professor in the Department of Computer Science, S.U.S. Govt. College, Matak Majri, Indri (Karnal). She has published a number of research papers in the various national and international journals. She has presented her research in various conferences. She has done her Ph.D in Computer Science and Applications from Department of Computer Science and Applications Kurukshetra University, Kurukshetra in the field of Computer Networks. Her research areas include Computer Networking, Machine Learning, Soft Computing, etc. She is having more than 16 years of teaching experience and more than 10 years of research experience. Email: aggshamit@gmail.com