A PROJECT REPORT

on

"Parkinson Disease Classification Using Ensemble and Quantum Neural Network Approaches"

Submitted to

KIIT Deemed to be University

In Partial Fulfilment of the Requirement for the Award of

BACHELOR'S DEGREE IN
COMPUTER SCIENCE & ENGINEERING

BY

Ayush Kumar	2105955
Aditya Srivastava	21051111
Dhrubojyoti Mahato	21051561
Harshita Anmol	21051565
Abhishek Kumar	21051620
Soham Sanyal	21051690

UNDER THE GUIDANCE OF
Dr. Mukesh Kumar

SCHOOL OF COMPUTER ENGINEERING
KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY
BHUBANESWAR, ODISHA - 751024

April 2025

KIIT Deemed to be University

School of Computer Engineering
Bhubaneswar, ODISHA 751024

CERTIFICATE

This is certify that the project entitled

"Parkinson Disease Classification Using Ensemble and Quantum Neural Network Approaches "

submitted by

Ayush Kumar	2105955
Aditya Srivastava	21051111
Dhrubojyoti Mahato	21051561
Harshita Anmol	21051565
Abhishek Kumar	21051620
Soham Sanyal	21051690

is a record of bonafide work carried out by them, in the partial fulfilment of the requirement for the award of Degree of Bachelor of Engineering (Computer Science & Engineering) at KIIT Deemed to be university, Bhubaneswar. This work is done during year 2025-2026, under our guidance.

Date: 09/04/2025

Mukesh Kumar 10/4/2025

Dr. Mukesh Kumar

Project Guide

Acknowledgements

We are profoundly grateful to **Dr. Mukesh Kumar** for his expert guidance and continuous encouragement throughout to see that this project rights its target since its commencement to its completion. ....

Ayush Kumar
Aditya Srivastava
Dhrubojyoti Mahato
Harshita Anmol
Abhishek Kumar
Soham Sanyal

ABSTRACT

With the use of machine learning algorithms and sophisticated data preprocessing techniques, this research presents a comprehensive system for the detection and classification of Parkinson diseases with high accuracy. The system is built to handle issues including missing values, unbalanced data, and feature variability while processing clinical information effectively. The model uses patient data to generate reliable predictions for a range of diseases by utilizing the power of machine learning classifiers such as Random Forest, Gradient Boosting, and Logistic Regression in conjunction with methods like data standardization and cross-validation.

The suggested method exhibits the capacity to efficiently forecast results by analyzing health-related characteristics like blood pressure, insulin levels, glucose levels, and BMI. The study illustrates the advantages and disadvantages of disease classification strategy by comparing and contrasting several models. The system's usefulness extends to health-care applications, such as enhancing diagnostic precision, recommending individualized treatments, and detecting disease early. This research lays the groundwork for future developments in automated health-care systems by using cutting-edge algorithms and data processing techniques, improving the effectiveness and precision of medical diagnostics.

Contents

Chapter 1
Introduction

Predicting medical conditions is an important aspect of health-care that aims to improve patient outcomes, schedule treatments, and diagnose patients early. However, because of the intricate interactions between medical variables, high- dimensional data, and noise in clinical records, disease prediction is frequently difficult. Conventional diagnostic techniques may not adequately capture the complex patterns found in patient data-sets since they mainly rely on statistical models and domain expertise. Advanced analytical methods are needed to manage the complexity and increase diagnosis accuracy as health-care systems produce data-sets that are larger and more varied.

Because it can analyse large data-sets, uncover hidden patterns, and assist in decision-making, machine learning (ML) has become a game-changing tool in medical diagnosis. Personalized medicine, cardiovascular risk assessment, and cancer detection are just a few of the fields that have used machine learning algorithms. Although traditional methods like support vector machines and decision trees have demonstrated potential, there is a rising need to investigate new ways that can address the drawbacks of traditional systems, like their limited scalability and difficulties in learning non-linear correlations.

This study looks into cutting-edge machine learning methods and how they can be used with quantum computing to improve disease prediction and classification. The following are the study's main goals:Comparing the performance of several data-sets: assessing the generalisability and accuracy of predictive models by applying them to a variety of medical data-sets.

1. Investigating different data splitting ratios: To improve learning and validation procedures, evaluate how training-to-testing data splits affect model performance.
2. Examining quantum neural network and ensemble methods: Quantum Convolution Neural Networks (QCNNs) and classical ensemble approaches are combined to investigate their potential benefits and synergy in processing high-dimensional, entangled data.

But problems still exist, such as:

1. Addressing missing values in important medical characteristics without sacrificing precision.
2. To prevent model bias, data-sets with unequal class distributions should be balanced.
3. By tackling these goals, the study hopes to shed light on how hybrid classical- quantum methods might improve medical diagnostics' predictive capabilities. The report also emphasises how quantum machine learning is transforming health-care analytics by making diagnostic tools more precise, scalable, and effective.

4. To prevent model bias, data-sets with unequal class distributions should be balanced.
   By tackling these goals, the study hopes to shed light on how hybrid classical- quantum methods might improve medical diagnostics' predictive capabilities. The report also emphasizes how quantum machine learning is transforming health-care analytics by making diagnostic tools more precise, scalable, and effective.

Chapter 2

Basic Concepts/ Literature Review

2.1 Current State of Machine Learning in Medical Diagnosis .

Medical diagnosis has been transformed by machine learning (ML), which offers precise and automated disease detection. In tasks including image classification, disease prediction, and prognosis, methods like support vector machines (SVM), decision trees, and deep learning models (such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs)) have shown excellent success rates. For example:

In medical imaging specialties like radiology, cardiology, CNNs have demonstrated exceptional accuracy.

RNNs are widely employed for temporal data, including patient history- based prediction and ECG signal analysis.

Nevertheless, there are still problems, including poor generalization between data-sets, data imbalance, and interpretability problems. Integrating multi-modal data from several medical fields is another issue.

2.2 Previous Works on Ensemble Methods.

By integrating predictions from several models, ensemble approaches increase the resilience of the model. Important methods include stacking, boosting, and bagging (e.g., Random Forest, AdaBoost, XGBoost). Ensemble approaches are frequently employed in medical diagnostics for:

- **Feature aggregation.** Combining features from diverse datasets to increase diagnostic precision.
- **Error reduction:** Improving dependability by addressing biases and variance in individual classifiers.

In recent research, ensemble methods have been included into deep learning architectures. For example, hybrid models that mix CNN outputs with Random Forests for medical picture categorization have been developed. Computational complexity and expense, however, continue to be major obstacles.

2.3 Overview of Quantum Neural Networks in Classification Tasks.

Quantum neural networks (QNNs) rapidly process complicated data by utilizing quantum mechanical concepts like entanglement and superposition. In applications related to medicine: QNNs have demonstrated promise in processing high-dimensional genomic, MRI, and CT data.

For tasks like multi-modal medical data fusion, where classical machine learning techniques fall short, hybrid quantum-classic models have been investigated.

According to studies, QNNs can perform faster and more scalable than classical approaches, particularly when dealing with enormous data-sets. However, because of hardware constraints such quantum decoherence and error correction problems, their actual implementation is still in its infancy.

2.4 Gaps in Existing Research.

- **Generalization across data**: Numerous models now in use are data set- specific and perform poorly when used on a variety of datasets. This restriction is addressed by your emphasis on multi-dataset integration, which offers a scalable approach to the classification of generalizable medical conditions.

- **Combining Quantum and Classical Methods**: Although ensemble approaches and QNNs have been studied separately, little is known about their combined potential. By examining hybrid models that capitalize on the advantages of both strategies, your project closes this gap.

- **Effective Computational Structures: -** Existing QNN implementations are difficult to develop with limited quantum technology and frequently demand large amounts of computational power. By suggesting innovative frameworks, your research may offer useful advice for effectively implementing QNNs.

- **Optimization of the Ensemble in Medical Contexts**: The majority of ensemble research concentrate on generic datasets, ignoring the particular difficulties associated with medical data, such as the need for interpretability and an unbalanced class distribution. Your strategy specifically addresses these issues.

Chapter 3 Methodology

Introduction:-

In order to prepare raw data for model training and evaluation, data preprocessing is an essential step in the machine learning workflow. To guarantee data quality and model fidelity, the preprocessing techniques used in this investigation are described in the steps that follow.

3.1 Data Cleaning -

- ❖ Missing Values: Utilizing [a particular technique, such as mean imputation for numerical variables and mode imputation for categorical variables], missing values were found and fixed.
- ❖ Approximately 5% of the data contained missing entries, which were imputed using median values for numerical features to minimize bias.
- ❖ Duplicate Records: To guarantee data consistency, duplicate entries were identified using the `duplicated()` method and then eliminated.

The interquartile range (IQR) approach and other statistical techniques were used to identify outliers, and values were capped at acceptable thresholds to handle them.

3.2 Feature Engineering -

- ❖ Categorical Variable Encoding: Use of [e.g., one-hot encoding, label encoding] was used to encode categorical features. Compatibility with the machine learning techniques was therefore guaranteed. The categorical variable `Region` was one-hot encoded into three distinct binary columns.
- ❖ Feature Scaling : To ensure that features contribute evenly to the model, the data distribution was normalized by standardizing numerical variables using StandardScaler."Standardization transformed the data to have a mean of 0 and a standard deviation of 1, facilitating gradient-based optimization.
- ❖ Feature Selection : Features were chosen according to their domain relevance and association with the target variable. To lower dimensionality, methods like recursive feature elimination (RFE) were used.

3.3 Handling Imbalanced Data -

- ❖ The Synthetic Minority Oversampling Technique (SMOTE) was used to oversample in order to rectify the class imbalance in the target variable.
- ❖ This enhanced model performance on minority classes and guaranteed a balanced distribution across classes.

3.4 Addressing Data Leakage -

- ❖ Every preprocessing operation was limited to the training set, and after fitting on the training data, alterations were applied to the test set. This prevented data leaks and guaranteed impartial assessment.

❖ Handling Imbalanced Data -

- The Synthetic Minority Oversampling Technique (SMOTE) was used to oversample in order to rectify the class imbalance in the target variable.

- This enhanced model performance on minority classes and guaranteed a balanced distribution across classes.

❖ Addressing Data Leakage -

- Every preprocessing operation was limited to the training set, and after fitting on the training data, alterations were applied to the test set. This prevented data leaks and guaranteed impartial assessment.

- Scaling parameters were computed on the training set and applied to the test set during evaluation.

Chapter 4

Experimental Setup

The experimental setup describes the procedures and settings used in the study to assess model performance and use cutting-edge methods, such as quantum neural networks and ensemble learning.

4.1 Data Splitting Ratios:-

Several training-to-testing ratios were investigated in order to look at how data splitting affected model performance. These comprised:
The following ratios are used for training: 90:10, 80:20, 70:30, 60:40, and 50:50. Goal:
Evaluate the classifiers' resilience to different training data sets.
Assess the consistency of performance, especially on smaller training sets.
To preserve class distribution within the training and test sets, stratified sampling was used for each split.

4.2 Classifiers Used in Ensemble Methods :-

To increase diversity and improve prediction accuracy, ensemble learning techniques used the following basis classifiers:

- ❖ Logistic Regression
- ❖ Random Forest Classifier
- ❖ Gradient Boosting Classifier
- ❖ Support Vector Classifier (SVC)
- ❖ Naïve Bayes (GaussianNB)
- ❖ k-Nearest Neighbors (KNN)

To take use of these classifiers' complimentary qualities, ensemble approaches were used to combine them.

4.3 Voting Techniques :-

- ❖ Voting Classifier: - To combine predictions from several classifiers, hard and soft voting techniques were used:
- ❖ Hard Voting: Determined by the ensemble's prediction of the majority class.
- ❖ Soft Voting: The class with the highest weighted probability was chosen by combining the probabilities from each classifier.
- ❖ To highlight better-performing models, weights for individual classifiers were adjusted according to how well they performed on their own. Because of their superior individual accuracies, logistic regression and gradient boosting were given higher weights in soft voting.

4.4 Stacking Techniques:-

- ❖ Stacking Ensemble:

-To find the best combinations of each model's outputs, a meta-model (logistic regression) was trained using basic classifier predictions.

-Random Forest, Gradient Boosting, SVC, and KNN were the base classifiers that were employed.
-Every layer of the stacking ensemble employed cross-validation to guard against overfitting and guarantee reliable generalization.

- ❖ By utilizing complimentary decision boundaries acquired by various classifiers, stacking enhanced performance.

4.5 Quantum Neural Network Implementation:

- ❖ To investigate the possibility of improving classification problems, quantum neural networks (QNNs) were put into practice using the TensorFlow Quantum module. TensorFlow Quantum is the implementation framework (TFQ).

- ❖ Quantum Circuit Design: -

Parameterized quantum circuits were used to encode feature vectors into quantum states.
Quantum gates were optimized for classification using variational quantum circuits (VQCs).

- ❖ In a hybrid quantum-classical architecture, intermediate outputs from the quantum layer are processed by a classical feedforward neural network to provide predictions.

Gradient descent was used to optimize both quantum and classical parameters simultaneously.

- ❖ Dataset Transformation: TFQ's feature encoding tools were used to normalize and map features into quantum states.

On the dataset, quantum neural networks performed competitively, with gains noted in some situations involving high-dimensional and unbalanced input.

Chapter 5

Model Development

5.1 Data Preprocessing and Preparation

5.1.1 Data Encoding and Transformation

To facilitate efficient machine learning analysis, medical diagnostic data frequently has to be carefully preprocessed. Diagnostic labels were encoded into a numerical format appropriate for computational analysis in order to tackle the binary classification challenge in this study. The following transformation was applied to the target variable:

- ❖ Samples that are malignant (M) are encoded as 1
- ❖ Samples that are benign (B) are encoded as 0.

The original diagnostic labels' semantic content is preserved but binary categorization is made easier using this encoding technique.

5.1.2 Feature Normalization Techniques

In order to lessen the effects of the different measurement scales and distributions present in medical datasets, feature normalization is essential. In order to handle data heterogeneity and enhance model performance, we applied a logarithmic adjustment to a few chosen characteristics.

Rationale for log transformation:

1. Reduce skewness in feature distributions
2. Compress the range of extreme values
3. Mitigate the impact of outliers
4. Normalize feature distributions to improve model learning

5.1.3 Feature Standardization

Standardization was implemented using StandardScaler to ensure consistent feature scaling across the dataset. This preprocessing step is crucial for:

- ❖ Ensuring equal feature contribution during model training
- ❖ Preventing features with larger magnitudes from dominating the learning process
- ❖ Improving performance of algorithms sensitive to feature scaling

Limitations of the parkinson disease are the two diffrentiators that can deine one’s disease whether there is any disease or not.

5.2 Machine Learning Model Development

5.2.1 Individual Classifier Architecture

We developed a diverse ensemble of machine learning classifiers to capture different aspects of the classification problem:

Logistic Regression

- ❖ Linear probabilistic classifier
- ❖ Optimal for binary classification problems
- ❖ Provides interpretable probability outputs
- ❖ Assumes linear relationship between features and log-odds

Random Forest Classifier

- ❖ Ensemble method based on decision tree architecture
- ❖ Handles complex, non-linear relationships
- ❖ Provides feature importance rankings
- ❖ Reduces overfitting through ensemble averaging

Gradient Boosting Classifier

- ❖ Sequential ensemble technique
- ❖ Builds models focusing on correcting previous models' errors
- ❖ Demonstrates high predictive performance across various domains

Support Vector Machine

- ❖ Identifies optimal hyperplane separating classes
- ❖ Effective in high-dimensional spaces
- ❖ Robust to complex decision boundaries

Additional Classifiers

- ❖ K-Nearest Neighbors
- ❖ Gaussian Naive Bayes

5.3 Model Evaluation Framework

Performance Metrics

The comprehensive evaluation strategy employed multiple performance indicators:

- ❖ Accuracy
- ❖ Precision
- ❖ Recall
- ❖ F1 Score
- ❖ Balanced Accuracy

- ❖ Receiver Operating Characteristic (ROC) Curve

Experimental Design

- ❖ Train-test split ratios: 10% to 50%
- ❖ Cross-validation techniques
- ❖ Visualization of model performance

5.4 Technological Implementation

Software and Libraries

- ❖ Programming Language: Python 3.x
- ❖ Core Libraries:
- ❖ Scikit-learn
- ❖ Pandas
- ❖ NumPy
- ❖ Advanced Libraries:
- ❖ TensorFlow
- ❖ TensorFlow Quantum
- ❖ Cirq

❖ Methodological Considerations

To create a strong framework for classifying medical conditions, the suggested methodology combines ensemble methods, quantum computing technologies, and classical machine learning techniques. In order to tackle the intricacy of medical diagnostic classification, the multifaceted approach:

1 Implementing rigorous data preprocessing
2 Utilizing diverse machine learning algorithms
3 Employing advanced ensemble and meta-learning techniques
4 Exploring quantum machine learning methodologies

Chapter 6
Results

6.1 Result Analysis for Classifiers Used

Using four datasets—Cancer, Parkinson, Diabetes, and Heart—the study assesses the performance of six classifiers: Logistic Regression (LR), Random Forest (RF), Gradient Boosting (GB), Support Vector Machine (SVM), k-Nearest Neighbors (KNN), and Naive Bayes (NB). For test sizes ranging from 0.1 to 0.5, key performance metrics including as accuracy, precision, recall, F1-score, and balanced accuracy were calculated, offering thorough insights into the efficiency and robustness of the model.

Model Performance in Datasets :

Parkinson Dataset: SVM and KNN outperformed ensemble techniques like RF and GB, demonstrating greater accuracy (0.950) and balanced accuracy (0.972). Metrics like precision and recall fell below 0.70 for the majority of test sizes, demonstrating Naive Bayes' poor performance and limited applicability to this dataset.

How Test Sizes Affect Performance

Smaller Test Sizes (0.1–0.3):

The metrics for the majority of models were stable and consistent across all datasets, with little decrease in performance. This implies that even with little testing data, these models function well.

Test sizes that are moderate (0.4):

Accuracy and F1-scores gradually declined, especially for ensemble approaches (GB, RF). SVM and LR, however, continued to have competitive metrics across datasets.

Greater test sizes (0.5) :

Resulted in more noticeable performance differences, with NB and KNN showing notable drops in recall and precision. On the other hand, SVM and RF continued to perform well, demonstrating their stability across different data splits.

2. Model Rankings and Comparisons Top Performers:

SVM continuously outperformed RF and LR in terms of accuracy, precision, and recall across all datasets. These models showed resilience and flexibility across a range of datasets and testing scenarios.

Underperformers:

Naive Bayes showed considerable limitations in handling more complex datasets like Parkinson and Diabetes, with a balanced accuracy dropping to as low as 0.361 for Parkinson.

Ensemble vs. Individual Models:

While ensemble methods (RF, GB) delivered consistent and competitive performance, simpler models like LR occasionally matched or outperformed them in accuracy and computational efficiency.

0.1

		Lr	Rf	Gb	Svm	Knn	Nb
Accuracy	Dataset 1	0.95	0.9	0.9	0.95	0.95	0.65
	Dataset 2	0.75	0.75	0.7	0.75	0.75	0.65
Precision	Dataset 1	0.95	0.9	0.9	0.95	0.96	0.78
	Dataset 2	0.71	0.56	0.55	0.56	0.73	0.54
Recall	Dataset 1	0.95	0.9	0.9	0.95	0.95	0.65
	Dataset 2	0.75	0.75	0.7	0.75	0.75	0.65
F1-Score	Dataset 1	0.94	0.9	0.9	0.94	0.95	0.70
	Dataset 2	0.71	0.64	0.62	0.64	0.74	0.59
Balanced Accuracy	Dataset 1	0.75	0.72	0.72	0.75	0.97	0.36
	Dataset 2	0.57	0.5	0.47	0.5	0.63	0.43

0.2

		Lr	Rf	Gb	Svm	Knn	Nb
Accuracy	Dataset 1	0.79	0.92	0.94	0.89	0.92	0.71
	Dataset 2	0.64	0.79	0.72	0.79	0.77	0.72
Precision	Dataset 1	0.90	0.92	0.95	0.90	0.92	0.78
	Dataset 2	0.64	0.63	0.62	0.63	0.74	0.62
Recall	Dataset 1	0.89	0.92	0.94	0.89	0.92	0.71
	Dataset 2	0.64	0.80	0.72	0.79	0.77	0.72
F1-Score	Dataset 1	0.88	0.92	0.94	0.88	0.92	0.74
	Dataset 2	0.64	0.70	0.66	0.70	0.75	0.66
Balanced Accuracy	Dataset 1	0.71	0.84	0.85	0.71	0.89	0.66
	Dataset 2	0.45	0.5	0.45	0.5	0.58	0.45

0.3

		Lr	Rf	Gb	Svm	Knn	Nb
Accuracy	Dataset 1	0.86	0.93	0.91	0.89	0.91	0.76
	Dataset 2	0.63	0.74	0.71	0.74	0.68	0.71
Precision	Dataset 1	0.86	0.93	0.91	0.89	0.91	0.81
	Dataset 2	0.56	0.56	0.62	0.56	0.59	0.62
Recall	Dataset 1	0.86	0.93	0.91	0.88	0.1	0.76
	Dataset 2	0.63	0.74	0.71	0.74	0.68	0.71
F1-Score	Dataset 1	0.85	0.92	0.91	0.86	0.91	0.77
	Dataset 2	0.59	0.64	0.64	0.64	0.62	0.64
Balanced Accuracy	Dataset 1	0.77	0.86	0.85	0.76	0.87	0.77
	Dataset 2	0.44	0.5	0.5	0.5	0.48	0.5

0.4

		Lr	Rf	Gb	Svm	Knn	Nb
Accuracy	**Dataset 1**	**0.85**	**0.92**	**0.89**	**0.87**	**0.94**	**0.73**
	Dataset 2	**0.64**	**0.73**	**0.69**	**0.73**	**0.71**	**0.71**
Precision	**Dataset 1**	**0.85**	**0.93**	**0.95**	**0.89**	**0.95**	**0.81**
	Dataset 2	**0.55**	**0.53**	**0.53**	**0.53**	**0.63**	**0.60**
Recall	**Dataset 1**	**0.85**	**0.92**	**0.89**	**0.87**	**0.94**	**0.73**
	Dataset 2	**0.64**	**0.73**	**0.69**	**0.73**	**0.70**	**0.70**
F1-Score	**Dataset 1**	**0.85**	**0.91**	**0.89**	**0.85**	**0.94**	**0.74**
	Dataset 2	**0.59**	**0.62**	**0.59**	**0.61**	**0.64**	**0.62**
Balanced Accuracy	**Dataset 1**	**0.79**	**0.85**	**0.84**	**0.75**	**0.90**	**0.77**
	Dataset 2	**0.45**	**0.5**	**0.47**	**0.5**	**0.51**	**0.49**

0.5

		Lr	Rf	Gb	Svm	Knn	Nb
Accuracy	**Dataset 1**	**0.85**	**0.92**	**0.88**	**0.89**	**0.85**	**0.73**
	Dataset 2	**0.70**	**0.73**	**0.71**	**0.75**	**0.75**	**0.65**
Precision	**Dataset 1**	**0.85**	**0.92**	**0.89**	**0.89**	**0.85**	**0.83**
	Dataset 2	**0.71**	**0.56**	**0.55**	**0.56**	**0.73**	**0.54**
Recall	**Dataset 1**	**0.85**	**0.92**	**0.88**	**0.89**	**0.85**	**0.73**
	Dataset 2	**0.75**	**0.75**	**0.7**	**0.75**	**0.75**	**0.65**
F1-Score	**Dataset 1**	**0.85**	**0.92**	**0.88**	**0.88**	**0.84**	**0.75**
	Dataset 2	**0.71**	**0.64**	**0.62**	**0.64**	**0.74**	**0.59**
Balanced Accuracy	**Dataset 1**	**0.80**	**0.87**	**0.85**	**0.78**	**0.75**	**0.78**
	Dataset 2	**0.57**	**0.5**	**0.47**	**0.5**	**0.63**	**0.43**

6.2 Result Analysis for Ensemble Methods

Using test size ratios of 0.1, 0.2, 0.3, 0.4, and 0.5, the effectiveness of several machine learning classifiers was assessed across several datasets, including those related to Parkinson's disease. Mean accuracy and standard deviation, which gauge each classifier's stability and predictive performance, were the metrics employed for assessment.

Parkinson Datasets :

The top performers : CatBoost (CAT), Extreme Gradient Boosting (XGB), and Extra Trees (ET), demonstrated consistently high mean accuracies across all test sizes. CatBoost maintained stability with a standard deviation of 0.012 and achieved the best accuracy of 0.975 (test size: 0.1).

Worst performer : The Decision Tree Classifier (DC), especially when test sizes were greater. Its accuracy was as low as 0.599 (test size: 0.5).

Trend: Due to the difficulty of having less training data, accuracy for the majority of classifiers declined significantly as test size increased.

Parkinson Dataset
Top Performers :CatBoost, Random Forest (RF), and Extra Trees all showed outstanding results. With a test size of 0.2, CatBoost had the greatest mean accuracy of 0.936.

Worst Performer: SGD had the highest variability and the lowest accuracy, falling to 0.547 (test size: 0.5).

Trend: The resilience of Extra Trees and CatBoost for this dataset was demonstrated by their consistent high performance across test sizes. The performance of other classifiers, including SGD and Gaussian Naive Bayes (GNB), varied significantly.

General Observation :

Overall Best Classifiers: XGB and CatBoost showed excellent stability and accuracy across test sizes and datasets. These models work especially well in situations that call for a high degree of resilience and dependability.
Basic Model Performance: Lower accuracy and greater variability were displayed by simpler models, including SGD and Decision Tree Classifier, suggesting their limitations when dealing with complex datasets.
Effect of Test Size: Most classifiers' accuracy decreased slightly with increasing test size, indicating that model performance may be impacted by smaller training datasets.

Conclusion:

Across a variety of datasets, CatBoost and XGB proved to be the most dependable classifiers, exhibiting consistent performance even with different test sizes. The findings emphasize how crucial it is to choose reliable classifiers in order to achieve high predicted accuracy and stability.

Dataset 1

Test Size	Models	Accuracy	Recall	F1 Score	BAccuracy
0.1	Stacking	0.9	0.94	0.94	0.72
	GBoost	0.9	0.94	0.94	0.72
	HGBoost	0.9	0.94	0.94	0.72
	CatBoost	0.9	0.94	0.94	0.72
0.2	Stacking	0.92	0.96	0.95	0.84
	GBoost	0.94	1	0.96	0.85
	HGBoost	0.92	0.96	0.95	0.84
	CatBoost	0.94	1	0.96	0.85
0.3	Stacking	0.93	1	0.95	0.86
	GBoost	0.91	0.97	0.94	0.85
	HGBoost	0.89	0.97	0.93	0.82
	CatBoost	0.94	1	0.96	0.9
0.4	Stacking	0.94	1	0.96	0.9
	GBoost	0.91	0.96	0.94	0.85
	HGBoost	0.93	0.98	0.95	0.89
	CatBoost	0.94	1	0.96	0.9

Dataset 2

Test Size	Models	Accuracy	Recall	F1 Score	B Accuracy
0.1	Stacking	0.75	1	0.85	0.5
	GBoost	0.7	0.93	0.82	0.46
	HGBoost	0.75	0.93	0.84	0.56
	CatBoost	0.75	1	0.85	0.5
0.2	Stacking	0.79	1	0.88	0.5
	GBoost	0.71	0.9	0.83	0.45
	HGBoost	0.56	0.70	0.72	0.35
	CatBoost	0.79	1	0.88	0.5
0.3	Stacking	0.74	1	0.88	0.45
	GBoost	0.71	0.93	0.82	0.49
	HGBoost	0.66	0.86	0.79	0.46
	CatBoost	0.74	1	0.85	0.5
0.4	Stacking	0.73	1	0.84	0.5
	GBoost	0.69	0.94	0.81	0.47
	HGBoost	0.67	0.91	0.80	0.47
	CatBoost	0.73	1	0.84	0.5

Tables For Ensemble Methods

6.3 Result analysis for Quantum Convolution Neural Network

6.3.1 Performance of Classifiers on Multiple Data-sets with Varying Test Sizes: Four datasets Cancer, Heart Disease, Diabetes, and Parkinson's disease-are used in this work to assess classifier performance at different test size ratios (0.1 to 0.5). To identify patterns in learning, the training and validation accuracies were documented for each test size ratio over a period of 20 epochs.

6.3.2 Parkinson's Dataset :

The Parkinson's dataset's performance revealed:

Training Accuracy:

With very slight gains over epochs, accuracy rates were low across all test size ratios. Stabilization for bigger test sizes was between 0.375 and 0.389.

Validation Accuracy: The accuracy of validation varied, reaching a maximum of

0.289 for test sizes of 0.1 but sharply decreasing for greater test sizes.

6.3.3 Conclusion

The results demonstrate the varying performance of classifiers across different datasets and test sizes. Key challenges include overfitting and generalization issues, with smaller test sizes yielding relatively better validation performance. Future work may focus on optimizing model architectures and employing techniques like cross-validation to improve generalization across datasets.

Underperformance of Simpler Models: Larger test sizes and complicated datasets were too much for simpler models like the Decision Tree Classifier (DTC) and Naive Bayes (NB). For the Parkinson dataset, for example, NB's balanced accuracy fell to 0.361, much below acceptable limits.

Comparing the Results with the State-of-the-Art: Advanced models like CatBoost and XGB are considered state-of-the-art classifiers for medical datasets because of their accuracy and stability, which are in line with recent research.

Chapter 7 Discussion

7.1 Interpretation of Results

According to the analysis:

Parkinson Dataset: SVM and k-Nearest Neighbors (KNN) performed the best with balanced accuracy of 0.972. Naive Bayes (NB),on the other hand, fared noticeably worse, underscoring its shortcomings when handling complicated datasets.

7.2 Comparative Analysis with Existing Methods

Comparing Ensemble vs. Individual Models: While ensemble techniques like Random Forest and Gradient Boosting (GB) produced reliable results, their computational efficiency occasionally fell short of that of more straightforward models like Logistic Regression.

CatBoost (CAT) and Extreme Gradient Boosting (XGB) are examples of advanced ensembles that have demonstrated their superior feature- handling abilities by repeatedly outperforming other ensembles.

Underperformance of Simpler Models: Larger test sizes and complicated datasets were too much for simpler models like the Decision Tree Classifier (DTC) and Naive Bayes (NB). For the Parkinson dataset, for example, NB's balanced accuracy fell to 0.361, much below acceptable limits.

Comparing the Results with the State-of-the-Art: Advanced models like CatBoost and XGB are considered state-of-the-art classifiers for medical datasets because of their accuracy and stability, which are in line with recent research.

7.3 Strengths and Limitations of the Approach

Strengths:

Generalizability was guaranteed by thorough study across several datasets (Parkinson).
A comprehensive performance evaluation was made possible by the use of multiple metrics, including accuracy, precision, recall, F1-score, and balanced accuracy.
Even in situations with little training data, ensemble techniques like CatBoost shown resilience.

Limitation:

Models such as Stochastic Gradient Descent (SGD) and Decision Trees performed inconsistently, especially when test numbers rose.
Naive Bayes and other simplistic models were unable to efficiently handle noisy or high-dimensional datasets.
The recall and precision limitations of a number of models, including KNN and NB, were revealed by larger test sizes (e.g., 0.5).

7.4 Perspectives from Various Data Ensemble Techniques and Ratios:

Test Size Effects:

Smaller test sizes (0.1–0.3): Models performed steadily and consistently across all datasets, with little loss in precision or accuracy.
Moderate test sizes (0.4): While SVM and LR maintained competitive metrics, accuracy started to decrease, especially for ensemble approaches like RF and GB.
Greater performance disparities were observed with larger test sizes (0.5), with models such as NB and KNN exhibiting notable declines in recall and precision. Both SVM and CatBoost showed stability and resilience.

Ensemble Techniques:

Top accuracy was regularly attained by CatBoost and XGB, outperforming other classifiers (e.g., 0.975 for Cancer with CatBoost at test size 0.1).
Compared to more complex models like XGB, simpler ensembles like RF and GB performed well but were more impacted by changes in test size.

7.5. Conclusion:

Summary of Key Findings:

Top Performers: Across all datasets, SVM, CatBoost, and XGB proved to be the most dependable classifiers, offering excellent accuracy, precision, and stability.

Underperformers: When dealing with complicated or high-dimensional datasets, Naive Bayes and Decision Tree classifiers showed notable limits.

Overall Findings: While ensemble approaches such as CatBoost shown greater robustness, the majority of classifiers saw modest performance drops as test numbers increased.

Advancements in the Field:

This study highlights how crucial it is to choose classifiers according to test size ratios and dataset features.
It demonstrates how sophisticated ensemble techniques like CatBoost and XGB can achieve high predicted accuracy and stability across a variety of datasets, validating their usefulness in the classification of medical conditions.

Suggestions for Upcoming Studies:

Hybrid Models: Create and assess hybrid strategies that enhance resilience by combining the advantages of individual and ensemble classifiers.

Extended Datasets: To further confirm generalizability, test the suggested techniques on a wider range of medical datasets.

Feature Engineering: To improve model performance, investigate more sophisticated feature engineering and selection techniques.

Computing Trade-offs: Examine the trade-offs between computing cost and performance in contexts with limited resources to gauge real-world applicability.

Methods of Deep Learning: Examine the differences between the best- performing classifiers in this work and deep learning models, especially for high-dimensional medical datasets.

8 References

1 Medical Classification Books Using Machine Learning Techniques: Books :

Christopher Bishop's Pattern Recognition and Machine Learning discusses important algorithms including SVMs and ensemble approaches.
Max Kuhn's Applied Predictive Modeling focuses on methods for classifying and evaluating models.

Documents:

R. C. Deo (2015). medical machine learning. Circulation: Talks about how machine learning is used in healthcare.
Guestrin, C., and Chen, T. (2016). A technical overview of XGBoost, a scalable tree boosting method.

2 UCI Machine Learning Repository for Medical Datasets:

Wisconsin Breast Cancer Dataset: Link PIMA Diabetes Dataset: Link Heart Disease Dataset:Link
Scikit-learn (sklearn.datasets.load_breast_cancer) provides the dataset.

3 Group Methods (CatBoost, XGBoost):

L. Prokhorenkova and associates (2018). CatBoost: category characteristics combined with unbiased boosting explains the benefits of CatBoost for managing category data.
Guestrin, C., and Chen, T. (2016). XGBoost is a tree boosting technique that is scalable. draws attention to its accuracy and scalability.

4 Model Assessment and Enhancement

Bengio, Y., and Bergstra, J. (2012). Grid search and random search can benefit from random search for hyper-parameter optimization.
D. Berrar (2019). Cross-validation: Describes methods of evaluation for small datasets

5 Tools and Documentation

Scikit-learn: https://scikit-learn.org/stable/documentation.html
CatBoost: https://catboost.ai/docs/
XGBoost: https://xgboost.readthedocs.io/en/stable

INDIVIDUAL CONTRIBUTION REPORT:

MULTI-DATASET Parkinson Disease Classification USING ENSEMBLE AND QUANTUM NEURAL NETWORK APPROACHES

Ayush Kumar

2105955

Abstract:

Multi-Dataset Parkinson Disease Classification Using Ensemble and Quantum Neural Network Approaches" is a project that attempts to create a solid foundation for precise and expandable medical diagnosis. This study aims to address issues with integrating heterogeneous medical datasets, boosting classification performance, and improving interpretability in diagnostic models by utilizing the strength of ensemble methods and the new potential of quantum neural networks (QNNs).

Individual contribution and findings:

Development of a Framework: conceived and created a revolutionary hybrid strategy that combines quantum neural networks (QNNs) with ensemble approaches.
outlined the hybrid system's design, striking a balance between quantum and conventional components for best results.

Individual contribution to project report preparation:

Structuring the Report: Drafted the overall structure of the report, ensuring a logical flow between sections such as the introduction, literature review, methodology, results, and conclusion.

Individual contribution for project presentation and demonstration:

Content Creation:
The presentation was planned and organized to successfully convey the goals, methods, findings, and conclusions of the project.Made presentations that illustrated difficult ideas including data integration procedures, ensemble approaches, and QNN architecture.

Mukesh Kr 18/4/2025

Ayush Kumar

Full Signature of Supervisor: ……………………………….

Full signature of the student: ………………………………..

INDIVIDUAL CONTRIBUTION REPORT:

MULTI-DATASET Parkinson Disease Classification USING ENSEMBLE AND QUANTUM NEURAL NETWORK APPROACHES

Aditya Srivastava
21051111

Abstract:

This project offers a strong foundation for predicting different diseases using cutting-edge data processing methods and machine learning. To estimate the probability of several diseases at once, the system combines clinical records, preprocesses data, and uses machine learning models. High accuracy and precision are attained by the model, opening the door for uses in early detection, personalised healthcare, and efficient illness management.

Individual contribution and findings:

My main responsibility was to build and implement the machine learning pipeline, which included choosing appropriate classification models, engineering features, and preparing the clinical information.

Important contributions consist of Creating a preprocessing strategy to encode categorical variables, handle missing values, and normalise data. Putting several models into practice and assessing how well they work, including XGBoost, Random Forest, and Logistic Regression. Experimenting with different data-splitting ratios in order to enhance accuracy by streamlining the testing and training process. I learnt a lot about feature engineering and model evaluation for practical applications through this project, and I also gained a lot of experience applying machine learning techniques to healthcare datasets.

Individual contribution to project report preparation:

My responsibility concentrated on the methodology part, was mostly my responsibility. This provided a thorough description of the feature engineering methods, machine learning models used, and data pretreatment procedures. I also made sure that the report was formatted correctly and that all references were cited.

Individual contribution for project presentation and demonstration:

With an emphasis on the approach and results parts, I was instrumental in the preparation and delivery of the project presentation. I presented the trained models' real-time predictions during the live demonstration and discussed the insights gleaned from performance metrics including accuracy, precision, recall, and F1-score.

10/4/2025

Full Signature of Supervisor: ……………………………….

Aditya Srivastava

Full signature of the student: ……………………………….

INDIVIDUAL CONTRIBUTION REPORT:

MULTI-DATASET Parkinson Disease Classification USING ENSEMBLE AND QUANTUM NEURAL NETWORK APPROACHES

Dhrubojyoti Mahato
21051561

Abstract:

Using Quantum Neural Networks (QCNN) and ensemble approaches, this project examines the classification of medical conditions across several datasets. For reliable generalization, the study places a strong emphasis on creating hybrid models, evaluating overfitting, and optimizing data ratios. My efforts included helping to build the model, partially coding, conceiving the study emphasis, and leading the preparation and presentation of the results.

Individual contribution and findings:

Contributed significantly to the examination and interpretation of QCNN results, covering issues with generalization, dataset-specific rformance patterns, and test size variations. helped with the hybrid QCNN models' coding, debugging, and hyperparameter adjustment for nhanced performance.compiled information on the effects of data ratios and suggested ensemble integration techniques to deal with performance instability and overfitting. highlighted the advantages and disadvantages of QCNN by providing a comparative analysis against other machine learning models

Individual contribution to project report preparation:

Authored critical parts hat included suggestions for further research, strengths and weaknesses, and an interpretation of the results. organized and improved the report, especially the discussion and conclusion arts, to guarantee cohesion and technical accuracy. helped put findings in context by offering insights on the influence of ensemble techniques and dataset features.

Individual contribution for project presentation and demonstration:

Oversaw the creation of the project presentation, which included performance data and important results visuals. presented the presentation's main points, focusing on the QCNN architecture, results analysis, and suggested enhancements. Model performance was demonstrated, and the practical implications of integrating QCNN with ensemble approaches were highlighted.

10/4/2025 | Dhrubojyoti Mahato

Full Signature of Supervisor: …………………………….

Full signature of the student: …………………………….

INDIVIDUAL CONTRIBUTION REPORT:

Multi-Dataset Parkinson Disease Classification Using Ensemble AND Quantum Neural Network Approaches

Harshita Anmol

21051565

Abstract:

Using Quantum Neural Networks (QCNN) and ensemble approaches, this project explores the classification of medical conditions across various datasets. The study emphasizes hybrid model development, overfitting analysis, and data ratio optimization to ensure reliable generalization.

Individual contribution and findings:

In the preprocessing phase, I transformed diagnostic labels into binary format, encoding malignant samples as 1 and benign samples as 0, to facilitate classification.While standardizing features with StandardScaler maintained uniform scaling, enhancing model performance and avoiding larger-magnitude features from dominating, logarithmic transformations were used to handle outliers, reduce skewness, and compress extreme values.
To handle both linear and non-linear data interactions, I put a number of machine learning techniques into practice. These included gradient boosting for sequential error correction, random forest for ensemble-based decision-making, and logistic regression for interpretability. In high-dimensional spaces, Support Vector Machines (SVM) maximized hyperplane separation, whereas Gaussian Naive Bayes and K-Nearest Neighbors improved model diversity. In order to incorporate cutting-edge techniques into the framework, I also investigated quantum approaches utilizing TensorFlow Quantum and Cirq.

Model evaluation involved experiments with varying train-test splits and cross- validation to ensure robustness. Metrics such as accuracy, precision, recall, F1 score, and ROC curve analysis provided a detailed assessment of performance, highlighting the strengths and weaknesses of individual classifiers and the ensemble.Additionally, I contributed significantly to the project report and presentation. I documented key methodologies and findings and delivered a presentation featuring the workflow, model architectures, and performance outcomes. This included visual aids and a live demonstration of real-time classification results.

Mukesh Ram 10/4/2025

Full Signature of Supervisor:
…………………………….

Harshita Anmol

Full signature of the student:
……………………………..

INDIVIDUAL CONTRIBUTION REPORT:

Multi-Dataset Parkinson Disease ClassificationUsing Ensemble and Quantum Neural Network Approaches

Abhishek Kumar

21051620

Abstract:

Using Quantum Neural Networks (QCNN) and ensemble approaches, this project explores the classification of medical conditions across various datasets. The study emphasizes hybrid model development, overfitting analysis, and data ratio optimization to ensure reliable generalization. Nishant Kumar significantly contributed to the conceptualization, coding, debugging, and evaluation of the project while assisting in report preparation and presentation.

Individual contribution and findings:

- Technical Contributions:

➢ Played a crucial role in coding and debugging the hybrid QCNN models, focusing on hyperparameter tuning to improve performance.
➢ Analyzed results related to QCNN's generalization, addressing issues such as dataset- specific performance patterns and variations in test set sizes.
➢ Compiled insights on data ratio effects, suggesting ensemble integration techniques to enhance stability and mitigate overfitting.

- Analytical Insights:

➢ Provided a comparative analysis of QCNN versus traditional machine learning models, highlighting the strengths and limitations of each.
➢ Explored the implications of combining ensemble techniques with QCNN to address challenges like performance variability.

Individual contribution to project report preparation:

Significant contributions were made to the project report by drafting key sections like future research suggestions, analyzing strengths and weaknesses, and interpreting findings. The discussion and conclusion were carefully organized for clarity and accuracy. Additionally, insights into how ensemble techniques and dataset characteristics affect model performance were provided, enhancing the overall quality of the report.

Individual contribution for project presentation and demonstration:

➢ Led the creation of the project presentation, including visualizations of performance data and key findings.
➢ Delivered major portions of the presentation, emphasizing the QCNN architecture, result analysis, and proposed improvements.
➢ Demonstrated the model's performance and discussed the practical implications of integrating QCNN with ensemble approaches.

18/4/2025

Abhishek Kumar

INDIVIDUAL CONTRIBUTION REPORT:

MULTI-DATASET Parkinson Disease ClassificationUSING ENSEMBLE AND QUANTUM NEURAL NETWORK APPROACHES

Soham Sanyal
21051690

Abstract:

As part of this study, I helped create a prediction model that uses input data to diagnose diseases. The algorithm forecasts the probability of four distinct diseases using machine learning techniques. I made sure the model's accuracy, precision, recall, and F1-score were all successfully assessed by carrying out a comprehensive outcome analysis. My research gave the team a better understanding of the performance measures and enabled them to make data-driven changes that improved the model's dependability.

Individual contribution and findings:

Analyzed the model's performance parameters, including confusion matrices, accuracy, precision, recall, and F1-scores, to conduct outcome analysis for each of the four disease predictions,highlighted patterns that needed to be adjusted in the training procedure or data preparation to identify places where the model performed poorly,Based on the results, methods for dataset augmentation and parameter adjustment were proposed to increase prediction accuracy, gave the team comprehensive statistical information to aid in decision-making, which directly enhanced the predictive model as a whole.

Overall, my contribution to the project involved thorough research, meticulous planning, and hands-on development of the pose estimation model using MediaPipe and OpenCV. Through this experience, I acquired valuable technical skills and insights into the practical implementation of machine learning techniques in real-world applications.

Individual contribution to project report preparation:

Created the project report's Result Analysis section, which included evaluations and performance indicators for every illness prediction.
To effectively communicate the data and findings, tables, graphs, and visual representations were created.
To ensure consistency and quality, I helped edit and proofread various parts of the report.

Individual contribution for project presentation and demonstration:

Created presentation slides pertaining to the result analysis, making sure they were clear and consistent with the main plot.
During the demonstration, the analytical results were explained, with a focus on the influence of result analysis on the model's optimization.

Mukesh Ku 18/4/2025

Soham Sanyal

Full Signature of Supervisor: ……………………………….

Full signature of the student: ………………………………..

ABSTRACT

ORIGINALITY REPORT

17%	12%	9%	11%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Georgia Institute of Technology Main Campus Student Paper	2%
2	Submitted to De Montfort University Student Paper	2%
3	Submitted to University of Hertfordshire Student Paper	1%
4	www.frontiersin.org Internet Source	1%
5	arxiv.org Internet Source	1%
6	O. P. Verma, Seema Verma, Thinagaran Perumal. "Advancement of Intelligent Computational Methods and Technologies - Proceedings of I International Conference on Advancement of Intelligent Computational Methods and Technologies (AICMT2023)", CRC Press, 2024 Publication	1%
7	Ashok Kumar, Geeta Sharma, Anil Sharma, Pooja Chopra, Punam Rattan. "Advances in Networks, Intelligence and Computing - International Conference on Networks, Intelligence and Computing (ICONIC-2023)", CRC Press, 2024 Publication	1%
	link.springer.com	

8	Internet Source	<1%
9	www.nature.com Internet Source	<1%
10	www.researchgate.net Internet Source	<1%
11	Houda Bichri, Adil Chergui, Mustapha Hain. "Investigating the Impact of Train / Test Split Ratio on the Performance of Pre-Trained Models with Custom Datasets", International Journal of Advanced Computer Science and Applications, 2024 Publication	<1%
12	journals.bilpubgroup.com Internet Source	<1%
13	Forrest Iandola, Kurt Keutzer. "Small neural nets are beautiful", Proceedings of the Twelfth IEEE/ACM/IFIP International Conference on Hardware/Software Codesign and System Synthesis Companion - CODES '17, 2017 Publication	<1%
14	H.L. Gururaj, Francesco Flammini, S. Srividhya, M.L. Chayadevi, Sheba Selvam. "Computer Science Engineering", CRC Press, 2024 Publication	<1%
15	ijrpr.com Internet Source	<1%
16	Arvind Dagur, Karan Singh, Pawan Singh Mehra, Dhirendra Kumar Shukla. "Artificial Intelligence, Blockchain, Computing and Security", CRC Press, 2023 Publication	<1%

17	www.medrxiv.org Internet Source	<1 %
18	Anu Sayal, Ashulekha Gupta, Janhvi Jha, N. Chaithra, Allagari Nithin. "chapter 6 Application of Artificial Intelligence in Healthcare Sector", IGI Global, 2024 Publication	<1 %
19	ebin.pub Internet Source	<1 %
20	pmc.ncbi.nlm.nih.gov Internet Source	<1 %
21	jazindia.com Internet Source	<1 %
22	www.mdpi.com Internet Source	<1 %
23	Submitted to BITS, Pilani-Dubai Student Paper	<1 %
24	Submitted to Central Queensland University Student Paper	<1 %
25	Submitted to Liberty University Student Paper	<1 %
26	"Data Science and Applications", Springer Science and Business Media LLC, 2024 Publication	<1 %
27	Submitted to University of Sunderland Student Paper	<1 %
28	edepot.wur.nl Internet Source	<1 %
29	uobrep.openrepository.com Internet Source	<1 %

30	H L Gururaj, Francesco Flammini, V Ravi Kumar, N S Prema. "Recent Trends in Healthcare Innovation", CRC Press, 2025 Publication	<1 %
31	www.diva-portal.se Internet Source	<1 %
32	Manuel Domínguez-Rodrigo, Gabriel Cifuentes-Alcobendas, Marina Vegara-Riquelme, Enrique Baquedano. "Reassessing Deep Learning (and Meta-Learning) Computer Vision as an efficient method to determine taphonomic agency in bone surface modifications", Cold Spring Harbor Laboratory, 2025 Publication	<1 %
33	iqac.pdpu.ac.in Internet Source	<1 %
34	www.fastercapital.com Internet Source	<1 %
35	Ankur Gupta. "Next Generation Computing and Information Systems", CRC Press, 2024 Publication	<1 %
36	Gaobo Liang, Lixin Zheng. "A transfer learning method with deep residual network for pediatric pneumonia diagnosis", Computer Methods and Programs in Biomedicine, 2020 Publication	<1 %
37	Sheikh Fahad Ahmad, Shadab Siddiqui, Rajib Debnath, Kakali Das, Figlu Mohanty, Sumit Hazra. "Advances in Computational Intelligence and its Applications - Navigating the Frontiers of AI, Machine Learning, and Data Science in the Digital Era International Conference on Advances in Computational	<1 %

Intelligence and its Applications (ICACIA-2023)", CRC Press, 2024
Publication

38	Usha Desai, Biswaranjan Acharya, Madhu Shukla, Varadraj Gurupur. "Decision Support System for Diabetes Healthcare: Advancements and Applications", River Publishers, 2025 Publication	<1 %
39	easychair.org Internet Source	<1 %
40	www.ijcseonline.isroset.org Internet Source	<1 %

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		