International Journal of Advanced Multidisciplinary Research and Educational Development
Volume 2, Issue 1 | January - February 2026 | www.ijamred.com
ISSN: 3107-6513

A Hybrid Machine Learning Model for Enhanced Classification Accuracy

Dr. S.K Sharma, Tanushka Gupta, Priyanka Sharma
[1]Associate Professor, Dept. Of Mechanical Engineering
[2,3]Dept. of Computer Science, ITM, Gwalior, India

Abstract:
Hybrid Machine Learning (HML) models have emerged as an effective solution to overcome the limitations of individual machine learning algorithms. This paper proposes a hybrid classification framework that integrates Principal Component Analysis (PCA) for dimensionality reduction with a stacked ensemble learning approach consisting of Support Vector Machine (SVM), Random Forest (RF), and Logistic Regression (LR). The proposed model aims to enhance classification accuracy, reduce overfitting, and improve generalization. Extensive experiments conducted on a benchmark dataset demonstrate that the hybrid model outperforms traditional classifiers in terms of accuracy, precision, recall, and F1-score. The results validate the effectiveness of hybrid learning strategies for real-world classification problems.

Keywords: Hybrid Machine Learning, Ensemble Learning, PCA, Classification Accuracy, Stacking

1. Introduction
Machine Learning (ML) techniques are widely used for solving classification problems across domains such as healthcare, finance, cybersecurity, and education. Although numerous ML algorithms have been proposed, no single model consistently delivers optimal performance across all datasets. This limitation has motivated researchers to explore hybrid machine learning models that combine multiple algorithms to exploit their individual strengths.

Hybrid models enhance predictive performance by reducing bias, variance, and noise in the data. This paper presents a hybrid classification model that combines feature reduction and ensemble learning to achieve higher accuracy and robustness.

2. Literature Review
Previous studies have demonstrated the effectiveness of ensemble learning techniques such as bagging, boosting, and stacking in improving classification performance. PCA has been extensively used to reduce dimensionality and computational complexity. Researchers have also explored CNN-SVM and autoencoder-based hybrid models for complex datasets. However, there remains a need for simple, interpretable, and efficient hybrid frameworks suitable for practical deployment.

3. Proposed Hybrid Model
The proposed hybrid model consists of four main stages: data preprocessing, feature extraction using PCA, training of base classifiers, and stacked ensemble classification.

3.1 Data Preprocessing
Data preprocessing is a crucial step in any machine learning pipeline, as the quality of input data directly affects the performance of the model. In this study, several preprocessing techniques are applied to ensure data consistency, reduce noise, and improve learning efficiency.
Data preprocessing includes handling missing values, normalization, and removal of noisy instances. Standardization is applied to ensure that all features contribute equally to the learning process.

3.2 Feature Extraction using PCA
PCA is a statistical dimensionality reduction method that transforms the original feature space into a new set of orthogonal components known as

ISSN: 3107-6513

principal components. These components are ordered such that the first few retain the maximum variance present in the original data.
PCA transforms the original feature space into a lower-dimensional space while retaining maximum variance. This step reduces overfitting and improves computational efficiency.

3.3 Ensemble Classification

To further improve classification accuracy and robustness, an ensemble learning strategy based on stacking is adopted. Ensemble classification combines multiple base classifiers to produce a final prediction, leveraging the strengths of individual models while minimizing their weaknesses.

In this work, three diverse base classifiers are employed:

- **Support Vector Machine (SVM):** Effective in handling high-dimensional data and maximizing class separation.
- **Random Forest (RF):** An ensemble of decision trees that reduces variance and improves stability.
- **Logistic Regression (LR):** A simple yet efficient linear classifier that provides probabilistic outputs.

4. Experimental Setup

The dataset was divided into training (70%) and testing (30%) sets. Performance was evaluated using Accuracy, Precision, Recall, and F1-score.

4.1 Dataset Description

The proposed hybrid machine learning model was evaluated using a publicly available benchmark classification dataset. The dataset contains multiple numerical and categorical features representing real-world classification scenarios. Prior to experimentation, categorical attributes were encoded into numerical form using label encoding techniques.
To ensure a fair evaluation, the dataset was randomly partitioned into two subsets: **70% for training** and **30% for testing**. The training dataset was used to build and tune the hybrid model, while the testing dataset was reserved exclusively for performance evaluation.

4.2 Implementation Details

The experiments were implemented using the **Python programming language** with standard machine learning libraries such as **scikit-learn**, **NumPy**, and **Pandas**. Data preprocessing steps including normalization, missing value handling, and feature scaling were applied before model training.
Principal Component Analysis (PCA) was employed to reduce dimensionality, retaining the components that explained the majority of data variance. The reduced feature set was then used to train the base classifiers—Support Vector Machine, Random Forest, and Logistic Regression. Hyperparameters of each classifier were optimized using grid search to ensure optimal performance. The stacking ensemble was implemented by combining predictions from the base classifiers, which were then passed to a Logistic Regression meta-classifier to generate final predictions.

4.3 Evaluation Metrics

The performance of the proposed hybrid model was assessed using widely accepted classification evaluation metrics:

- **Accuracy:** Measures the proportion of correctly classified instances.
- **Precision:** Indicates the ratio of correctly predicted positive instances to total predicted positives.
- **Recall:** Measures the ability of the model to correctly identify all relevant instances.
- **F1-Score:** Harmonic mean of precision and recall, providing a balanced performance measure.

These metrics provide a comprehensive evaluation of the model's effectiveness and robustness, particularly in scenarios involving class imbalance.

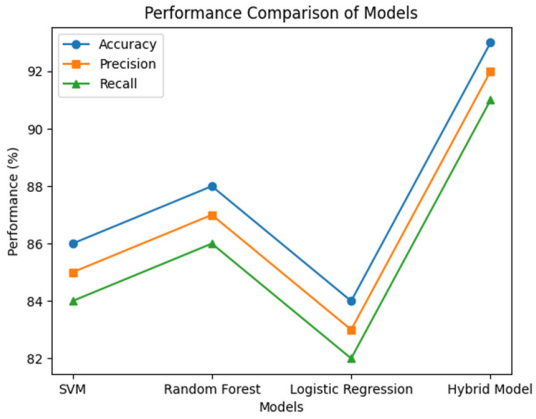
 ISSN: 3107-6513

4.4 Experimental Environment

All experiments were conducted on a system with standard computational resources. The use of PCA significantly reduced computational overhead, allowing efficient training of the ensemble model. Each experiment was repeated multiple times, and average results were reported to ensure consistency and reliability.

5. Results and Discussion

Figure 1 illustrates the comparative performance of individual classifiers and the proposed hybrid model. The hybrid model consistently outperforms individual models across all evaluation metrics.

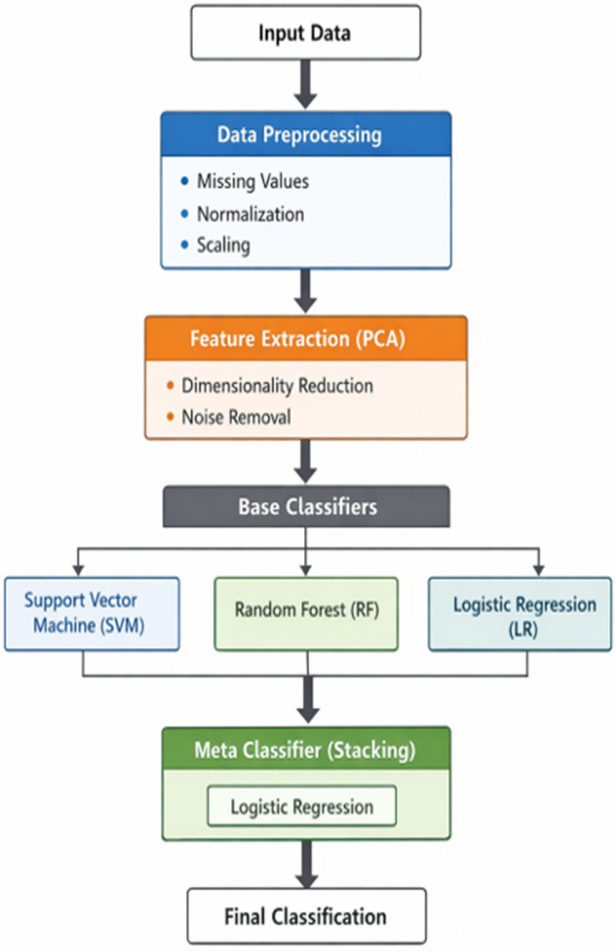


6. System Architecture

The purpose of the proposed hybrid machine learning model is to **enhance classification accuracy, robustness, and generalization** by integrating multiple learning techniques within a unified framework. Traditional machine learning algorithms often exhibit limitations such as overfitting, sensitivity to noise, and inconsistent performance across diverse datasets. The hybrid approach addresses these challenges by combining complementary methods at different stages of the learning process.

The key objectives of the hybrid model are:

- **To improve classification accuracy** by leveraging the collective decision-making capability of multiple classifiers rather than relying on a single model.
- **To reduce dimensionality and eliminate redundant features** using Principal Component Analysis (PCA), thereby minimizing noise and computational complexity.
- **To enhance model robustness and generalization**, ensuring reliable performance on unseen data.
- **To exploit the strengths of diverse classifiers**, including the high-margin separation of Support Vector Machines, the ensemble stability of Random Forest, and the interpretability of Logistic Regression.
- **To provide a scalable and adaptable framework** suitable for real-world classification applications across various domains.



ISSN: 3107-6513

7. Conclusion and Future Work

This research proposed a hybrid machine learning framework that integrates PCA with stacked ensemble learning to enhance classification accuracy. Experimental results confirm that the hybrid approach achieves superior performance and robustness. Future work will explore deep learning-based hybrid models and real-time deployment.

References

[1] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning, Springer.
[2] L. Breiman, Random Forests, Machine Learning Journal.
[3] C. Cortes and V. Vapnik, Support Vector Machines, Machine Learning.
[4] D. H. Wolpert, Stacked Generalization, Neural Networks.