ISSN: 2581-3404 (Online) **IF: 5.445 (SJIF)**
International Journal of Innovative Research in Technology and Management, Volume-6, Issue-4, 2022.

Application of Artificial Intelligence Techniques for Data Classification and Decision Support Systems

Navneet Verma[1], Deepak Singla[2], Vivek[3], Sorabh Gupta[4]
Department of CSE [1]
Department of CSE (AI & ML) [2]
Department of Information Technology[3,4]
Panipat Institute of Engineering and Technology, Panipat, India [1,2,3,4]

Abstract: *Data classification is the core step in modern decision support systems (DSS), yet practitioners often face the question of which Artificial Intelligence (AI) technique to trust and when an automated decision should be handed to a human expert. This paper presents an experimental study of seven AI classifiers—Naive Bayes, Decision Tree, k-Nearest Neighbours, Logistic Regression, Support Vector Machine, Random Forest and a Multi-Layer Perceptron—on three benchmark datasets (Iris, Wine and Breast Cancer Wisconsin) using stratified 10-fold cross-validation. A confidence-based decision support layer is then added that accepts a prediction automatically only if its class probability exceeds a threshold and otherwise refers the case to an expert. Results show that the Support Vector Machine achieved the best mean accuracy (97.11%) across datasets, while the Decision Tree was the weakest (90.70%). On the Breast Cancer data, the decision layer raised the accuracy of a Random Forest from 95.61% to 99.34% on automatically decided cases while referring 20.56% of cases to an expert. The findings show that combining accurate classifiers with an uncertainty-aware decision layer gives a practical and safer decision support design.*

Keywords: Artificial intelligence, data classification, decision support system, machine learning, cross-validation, selective classification.

Introduction

Organisations in healthcare, finance, agriculture and security now generate far more data than people can inspect manually. Decision support systems (DSS) use computational models to turn such data into recommendations, and Artificial Intelligence (AI) has become the main engine behind them [4], [18]. At the heart of most DSS is data classification: assigning each record to one of several predefined classes, such as benign or malignant, genuine or fraudulent, safe or malicious.

Deep learning has produced striking results in image and language tasks [18], [19], [22], [14], [9], and has reached expert-level performance in medical diagnosis [8], [17]. However, for the structured, tabular data that dominates real decision support, classical learners remain highly competitive. A large-scale comparison of 179 classifiers on 121 datasets found that Random Forests and SVMs were among the best performers [21]. Practitioners therefore need clear, reproducible evidence of how different techniques behave on the same data.

A second concern is trust. A DSS that always outputs an answer, even for ambiguous cases, can be dangerous in sensitive fields such as medicine [3]. Explainability and uncertainty handling are therefore treated as central requirements of responsible AI [1], [2], [6], [11]. Selective classification, where a model abstains when it is not confident, offers a simple way to improve reliability [12].

The contributions of this paper are: (i) a controlled comparison of seven AI classifiers on three benchmark datasets using the same pre-processing and validation protocol; (ii) a confidence-based decision support layer

that trades coverage for accuracy; and (iii) a complete set of measured results, including accuracy, precision, recall, F1-score, a confusion matrix and a coverage–accuracy analysis.

Related Work

Early deep networks showed that learned representations can outperform hand-crafted features [22], and later architectures such as residual networks made very deep models trainable [14]. The attention mechanism further changed sequence modelling [9], and the Adam optimiser became a default training method [20]. A broad overview of this progress is given in [19]. Gradient-boosted trees, as implemented in XGBoost, remain a strong choice for tabular data [15].

In applied decision making, AI has been used to support managerial and clinical choices, but the literature stresses the need for human oversight and transparency [4], [3]. Methods such as LIME [16] and SHAP [10] explain individual predictions, and surveys classify the available explanation techniques [6], [2], [1]. Calibration studies show that modern neural networks are often over-confident [13], which affects any confidence-based rule. Proper evaluation, particularly cross-validation, is discussed in [7]. Our work combines these ideas in a small, reproducible framework.

Methodology

A. System Overview

The proposed workflow is shown in Fig. 1. Raw data are cleaned and standardised, split by stratified 10-fold cross-validation, and passed to seven classifiers. The output probabilities are then processed by the decision support layer, which either issues an automatic decision or refers the case to a human expert.

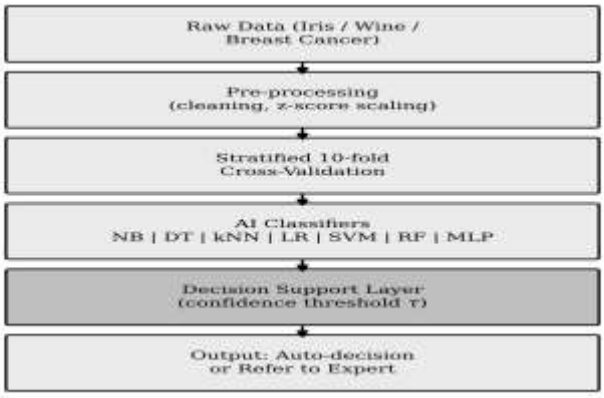


Fig. 1. Architecture of the proposed classification and decision support workflow.

B. Datasets

Three benchmark datasets from the UCI repository [5] were used (Table I). They differ in size, number of features and number of classes, which allows classifier behaviour to be compared under different conditions.

TABLE I. DATASETS USED

Dataset	Samples	Features	Classes
Iris	150	4	3
Wine	178	13	3
Breast Cancer	569	30	2

C. Classifiers

Seven techniques were selected to cover probabilistic, tree-based, instance-based, linear, kernel, ensemble and neural approaches: Gaussian Naive Bayes (NB), Decision Tree (DT), k-Nearest Neighbours with k = 5 (kNN), Logistic Regression (LR), Support Vector Machine with RBF kernel (SVM), Random Forest with 200 trees (RF), and a Multi-Layer Perceptron with two hidden layers of 32 and 16 neurons (MLP). Distance- and gradient-based models (kNN, LR, SVM, MLP) used z-score standardisation inside a pipeline so that no information leaked from test folds.

D. Evaluation Protocol

Each model was evaluated with stratified 10-fold cross-validation (random seed 42) [7]. Accuracy, macro-averaged precision, macro-averaged recall and macro-averaged F1-score were computed per fold and averaged. The standard deviation (SD) of accuracy across folds is also reported. Experiments were implemented in Python using the scikit-learn library.

E. Decision Support Layer

For a case x, a classifier outputs class probabilities p(c|x). The decision layer takes the highest probability, called the confidence, and applies the rule: if max $p(c|x) \geq \tau$, the predicted class is accepted automatically; otherwise the case is referred to a human expert. This is a form of selective classification [12]. Two quantities are measured: coverage, the fraction of cases decided automatically, and selective accuracy, the accuracy on those cases. The experiment was run on the Breast Cancer data with Random Forest probabilities obtained from out-of-fold predictions.

Results and Discussion

A. Classifier Comparison

Table II lists the measured results for every classifier and dataset. Fig. 2 summarises accuracy.

TABLE II. 10-FOLD CROSS-VALIDATION RESULTS (%)

Classifier	Acc.	Prec.	Rec.	F1	SD
Iris					
Naive Bayes	95.33	96.05	95.33	95.26	5.21
Decision Tree	93.33	94.54	93.33	93.20	5.16
kNN (k=5)	96.00	96.98	96.00	95.88	5.33
Logistic Reg.	95.33	96.05	95.33	95.26	5.21
SVM (RBF)	96.00	96.44	96.00	95.97	4.42
Random Forest	95.33	96.05	95.33	95.26	5.21
MLP (ANN)	94.00	95.21	94.00	93.93	6.96
Wine					
Naive Bayes	97.78	98.00	97.94	97.83	3.69
Decision Tree	87.19	88.62	87.24	87.18	7.44
kNN (k=5)	97.22	97.58	97.62	97.29	4.48
Logistic Reg.	98.33	98.47	98.38	98.29	2.55
SVM (RBF)	97.78	98.13	97.63	97.73	3.69
Random Forest	98.89	98.97	99.05	98.93	2.22

MLP (ANN)	98.33	98.55	98.38	98.34	2.55
Breast Cancer					
Naive Bayes	93.68	94.13	92.63	93.08	3.05
Decision Tree	91.57	91.41	90.86	90.95	2.04
kNN (k=5)	97.01	97.54	96.17	96.74	1.76
Logistic Reg.	97.54	97.78	97.07	97.33	1.95
SVM (RBF)	97.54	97.60	97.27	97.36	1.95
Random Forest	95.61	95.64	95.04	95.25	2.86
MLP (ANN)	97.37	97.52	96.93	97.14	2.39

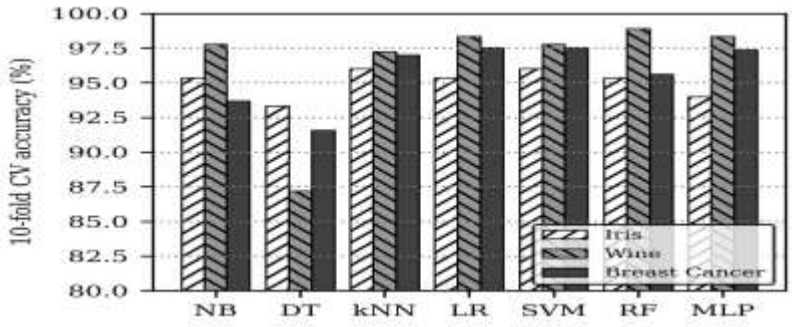


Fig. 2. Cross-validated accuracy of the seven classifiers on the three datasets (NB: Naive Bayes, DT: Decision Tree, kNN, LR: Logistic Regression, SVM, RF: Random Forest, MLP).

On Iris, kNN and SVM were best (96.00%), though differences between most models were within one standard deviation. On Wine, Random Forest achieved the highest accuracy (98.89%) and the lowest variance, while the Decision Tree dropped to 87.19%. On Breast Cancer, Logistic Regression and SVM tied at 97.54%, followed closely by the MLP (97.37%) and kNN (97.01%).

Averaged over the three datasets, the mean accuracies were: SVM 97.11%, LR 97.07%, kNN 96.74%, RF 96.61%, MLP 96.57%, NB 95.60% and DT 90.70%. A single Decision Tree is therefore consistently weaker than ensemble, kernel and linear methods, in line with the observation that no single method dominates every problem but a few families are reliably strong [21]. Notably, simple models such as Logistic Regression matched the neural network, which supports the view that deeper models are not always needed for small tabular data.

B. Error Analysis

Fig. 3 shows the confusion matrix of the Random Forest on the Breast Cancer dataset (out-of-fold predictions). Of 212 malignant cases, 197 were identified correctly and 15 were predicted as benign; of 357 benign cases, 347 were correct and 10 were flagged as malignant. In a medical DSS, the 15 missed malignant cases are the most costly errors, which motivates the decision layer below.

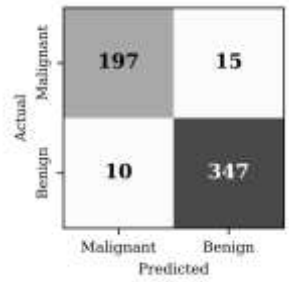


Fig. 3. Confusion matrix of Random Forest on the Breast Cancer dataset.

C. Effect of the Decision Support Layer

TABLE III. COVERAGE AND ACCURACY VERSUS THRESHOLD (RF, BREAST CANCER)

τ	Coverage (%)	Accuracy (%)	Referred	Auto
0.50	100.00	95.61	0	569
0.60	96.31	97.26	21	548
0.70	92.62	98.67	42	527
0.80	87.17	98.99	73	496
0.90	79.44	99.34	117	452
0.95	72.76	99.28	155	414

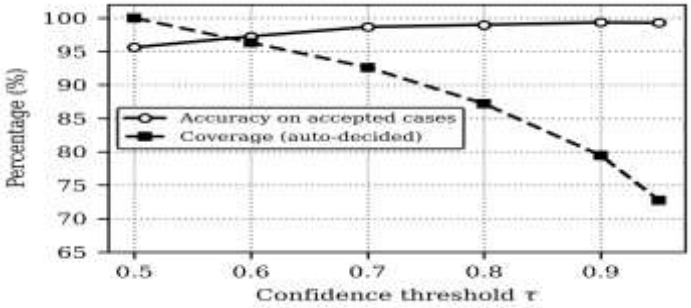


Fig. 4. Trade-off between automatic coverage and accuracy as the confidence threshold increases.

Table III and Fig. 4 show the trade-off. Without a threshold ($\tau = 0.50$) every case is decided automatically with 95.61% accuracy. Raising τ to 0.70 increases accuracy on automated cases to 98.67% while 42 cases (7.38%) go to an expert. At $\tau = 0.90$, the system reaches 99.34% accuracy on 79.44% of the cases, and 117 uncertain cases are referred. Beyond this point accuracy no longer improves (99.28% at $\tau = 0.95$) while the expert workload grows, so $\tau \approx 0.80$–0.90 is a sensible operating range for this dataset. The result confirms that selective classification can make AI decisions safer without changing the underlying model [12].

D. Discussion

Three lessons emerge. First, model choice matters less than data preparation and validation: scaled linear and kernel models were as good as an MLP. Second, a single decision tree is not recommended as the only engine of a DSS, although it is easy to interpret. Third, the decision layer gives a measurable safety benefit at very low cost. Its effectiveness depends on well-calibrated probabilities [13], so calibration should be checked in

deployment. For high-stakes use, model explanations [10], [16] should accompany each recommendation [1], [11].

E. Limitations

The datasets are small and clean, so results may not carry over to large, noisy, imbalanced or streaming data. Hyper-parameters were kept at reasonable defaults rather than tuned extensively, and only one decision rule (maximum probability) was tested. Statistical significance tests and cost-sensitive thresholds are left for future work.

Conclusion and Future Scope

This paper compared seven AI techniques for data classification and added a confidence-based decision support layer. SVM and Logistic Regression gave the best overall accuracy, Random Forest was best on Wine, and the Decision Tree was the weakest. The decision layer raised accuracy on automated Breast Cancer cases from 95.61% to 99.34% by referring about one fifth of the cases to experts. Future work will include larger real-world datasets, gradient-boosted and deep models [15], [14], calibration methods [13], explainability integration [10], [16] and cost-sensitive thresholds.

References

[1] W. Samek, G. Montavon, S. Lapuschkin, C. J. Anders, and K.-R. Müller, "Explaining deep neural networks and beyond: A review of methods and technologies," Proc. IEEE, vol. 109, no. 3, pp. 247–278, Mar. 2021.
[2] A. B. Arrieta et al., "Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," Inf. Fusion, vol. 58, pp. 82–115, Jun. 2020.
[3] A. Rajkomar, J. Dean, and I. Kohane, "Machine learning in medicine," N. Engl. J. Med., vol. 380, no. 14, pp. 1347–1358, Apr. 2019.
[4] Y. Duan, J. S. Edwards, and Y. K. Dwivedi, "Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda," Int. J. Inf. Manage., vol. 48, pp. 63–71, Oct. 2019.
[5] D. Dua and C. Graff, "UCI Machine Learning Repository," Univ. California, Irvine, School Inf. Comput. Sci., 2019. [Online]. Available: http://archive.ics.uci.edu/ml
[6] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi, "A survey of methods for explaining black box models," ACM Comput. Surv., vol. 51, no. 5, Art. no. 93, 2018.
[7] S. Raschka, "Model evaluation, model selection, and algorithm selection in machine learning," arXiv:1811.12808, 2018.
[8] A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, no. 7639, pp. 115–118, 2017.
[9] A. Vaswani et al., "Attention is all you need," in Proc. Adv. Neural Inf. Process. Syst. (NIPS), 2017, pp. 5998–6008.
[10] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Proc. Adv. Neural Inf. Process. Syst. (NIPS), 2017, pp. 4765–4774.
[11] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," arXiv:1702.08608, 2017.
[12] Y. Geifman and R. El-Yaniv, "Selective classification for deep neural networks," in Proc. Adv. Neural Inf. Process. Syst. (NIPS), 2017, pp. 4878–4887.
[13] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," in Proc. 34th Int. Conf. Mach. Learn. (ICML), 2017, pp. 1321–1330.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp. 770–778.
[15] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2016, pp. 785–794.
[16] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?': Explaining the predictions of any classifier," in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2016, pp. 1135–1144.
[17] V. Gulshan et al., "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," JAMA, vol. 316, no. 22, pp. 2402–2410, 2016.
[18] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015.
[19] J. Schmidhuber, "Deep learning in neural networks: An overview," Neural Netw., vol. 61, pp. 85–117, Jan. 2015.
[20] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in Proc. Int. Conf. Learn. Represent. (ICLR), 2015.
[21] M. Fernández-Delgado, E. Cernadas, S. Barro, and D. Amorim, "Do we need hundreds of classifiers to solve real world classification problems?," J. Mach. Learn. Res., vol. 15, pp. 3133–3181, 2014.
[22] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in Proc. Adv. Neural Inf. Process. Syst. (NIPS), 2012, pp. 1097–1105.