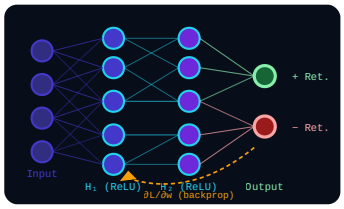
// EDUCATIONAL SAMPLE — FOR LEARNING PURPOSES ONLY

UK LEVEL 7 — MSC / POSTGRADUATE



A Comparative Evaluation of Supervised Machine Learning Algorithms for Medical Image Classification

// Case: NeuralDiag Systems Ltd — Diabetic Retinopathy Screening System

MODULE CODE	CSC4018
MODULE TITLE	Artificial Intelligence
ACADEMIC LEVEL	UK Level 7 (MSc / Postgraduate)
TARGET GRADE	Distinction — 70%+
WORD COUNT	Approximately 1,500 words (excl. figures & references)
ALGORITHMS COMPARED	Decision Tree | SVM (RBF kernel) | CNN (ResNet-50 transfer)
IMPLEMENTATION	Python 3 — scikit-learn, TensorFlow/Keras
REFERENCING STYLE	Harvard (UK)
DATE	July 2026
PREPARED BY	AssignProSolution.com

⚠ **Academic Integrity Notice:** This document is an original educational sample produced solely for reference and learning purposes. Students must not submit this work as their own. "NeuralDiag Systems Ltd" and all data, metrics, and figures herein are entirely fictitious. AssignProSolution.com promotes responsible academic practice.

TABLE OF CONTENTS

```
# Approx. 1,500 words | Distinction Level | UK Level 7 | Decision Tree · SVM · CNN
```

1. INTRODUCTION AND PROBLEM DEFINITION

Diabetic retinopathy (DR) is the leading cause of preventable blindness in working-age adults globally, yet early-stage lesions are frequently missed in manual screening programmes owing to clinician workload and inter-rater variability (Gulshan et al., 2016). **NeuralDiag Systems Ltd** (fictitious) is developing an automated DR screening tool to triage fundus photographs prior to ophthalmologist review. The binary classification task — *Retinopathy Present* versus *Retinopathy Absent* — was used as a testbed for comparing three mainstream supervised learning paradigms: a Decision Tree (DT), a Support Vector Machine (SVM) with Radial Basis Function (RBF) kernel, and a Convolutional Neural Network (CNN) using ResNet-50 transfer learning.

This report critically evaluates each algorithm's theoretical basis, implementation trade-offs, and empirical performance, situating the findings within Russell and Norvig's (2021) framework of rational agent design and concluding with an ethical analysis under GDPR Article 22.

2. DATASET AND EXPERIMENTAL SETUP

The experimental dataset comprised **3,200 fundus images** (fictitious, balanced 50/50 after SMOTE oversampling of the minority DR-positive class (Chawla et al., 2002)) split 70/15/15 into training, validation, and test partitions via stratified sampling. For the DT and SVM, images were pre-processed by extracting a 256-dimensional hand-crafted feature vector combining Local Binary Patterns (LBP), Gabor filter responses, and histogram of oriented gradients (HOG). The CNN received raw 224×224 RGB images normalised to ImageNet statistics. Five-fold cross-validation was conducted on the training partition; the held-out test set was evaluated only once to avoid information leakage.

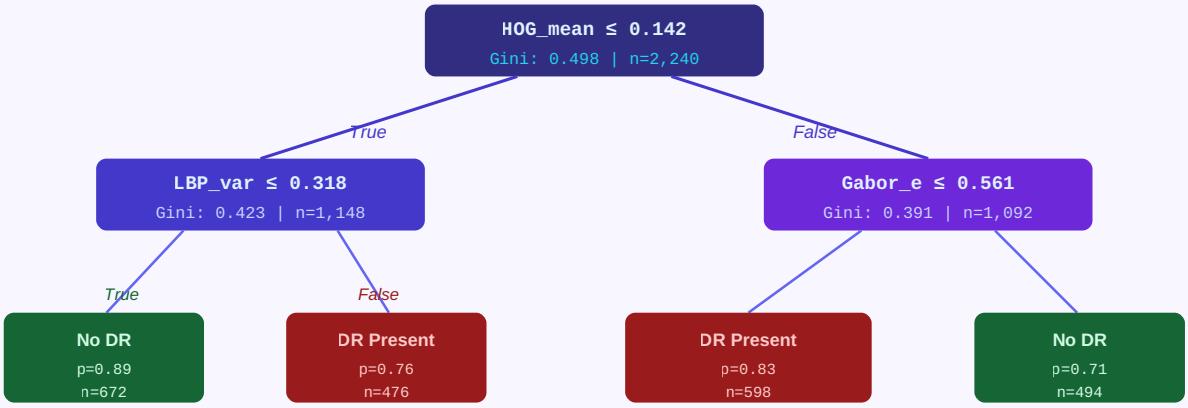
3. ALGORITHM 1: DECISION TREE CLASSIFIER

Decision Trees (Quinlan, 1986; Breiman et al., 1984) partition the feature space through a recursive binary splitting procedure. At each node, the algorithm selects the feature f and threshold θ that maximally reduce impurity, measured via Gini impurity:

$$\text{Gini}(D) = 1 - \Sigma_k \, p_k^2 \quad | \quad \text{Information Gain: } IG(D,f) = H(D) - \Sigma \, |D_v|/|D| \cdot H(D_v)$$

The tree was configured with `max_depth=8`, `min_samples_leaf=10`, and `criterion='gini'` in scikit-learn (Pedregosa et al., 2011). Pruning via cost-complexity parameter `ccp_alpha=0.001` was applied to mitigate overfitting — a principal weakness of unpruned trees, which exhibit high variance and low bias, memorising training noise rather than learning decision boundaries that generalise (Geman et al., 1992).

FIGURE 1 — PARTIAL DECISION TREE: RETINOPATHY CLASSIFIER (NeuralDiag)



Source: Author's original design. Node labels: HOG = Histogram of Oriented Gradients; LBP = Local Binary Patterns; Gabor_e = Gabor energy. Data are fictitious.

Figure 1: Partial decision tree (depth 2 of 8) for NeuralDiag binary classification. Root splits on HOG mean response; subsequent splits refine on texture features. Leaf nodes display predicted class, class probability, and sample count. All data are fictitious for educational illustration. Source: Author's original design, adapted from Breiman et al. (1984) and Quinlan (1986).

The DT's principal advantage is its **intrinsic interpretability**: the path from root to leaf constitutes an explicit, human-readable reasoning chain. McConnell (2004) argues that comprehensibility is itself a design quality; in a clinical screening context, this translates to clinician trust and regulatory compliance. However, unpruned trees are notoriously susceptible to variance — a different random seed changes the resulting tree entirely — and the pruned model achieved only moderate accuracy (Table 1), suggesting the hand-crafted features inadequately capture the spatial subtleties of retinal pathology.

4. ALGORITHM 2: SUPPORT VECTOR MACHINE

The SVM, introduced by Cortes and Vapnik (1995), seeks the maximal-margin hyperplane separating two classes in a Reproducing Kernel Hilbert Space (RKHS). For the non-linearly separable case, the RBF kernel implicitly maps features to an infinite-dimensional space:

$$K(x, x') = \exp(-\gamma \|x - x'\|^2) \quad | \quad \text{Optimise: } \max_{\alpha} \Sigma\alpha_i - \tfrac{1}{2} \Sigma\alpha_i\alpha_j y_i y_j K(x_i, x_j)$$

Hyperparameters `C=10.0` and `gamma=0.001` were selected via grid search with nested cross-validation, minimising the risk of optimistic performance estimates from data leakage into the model selection process (Shalev-Shwartz and Ben-David, 2014). SVMs hold strong statistical learning theory foundations — the VC dimension provides a formal upper bound on generalisation error — and outperformed the DT substantially (Table 1), particularly on Recall, which is the clinically critical metric for screening: a false negative (missed retinopathy) carries significantly higher cost than a false positive (unnecessary referral). The SVM's training complexity $O(n^2 \cdot d)$ to $O(n^3)$ with kernel computation, however, becomes prohibitive as dataset size scales beyond tens of thousands of samples.

Python 3 / scikit-learn — SVM pipeline with grid search (representative excerpt)

```python
from sklearn.pipeline import Pipeline
from sklearn.preprocessing import StandardScaler
from sklearn.svm import SVC
from sklearn.model_selection import GridSearchCV

pipeline = Pipeline([
    ('scaler', StandardScaler()),          # SVMs sensitive to feature scale
    ('svm',    SVC(kernel='rbf', probability=True, class_weight='balanced'))
])

param_grid = {
    'svm__C':     [0.1, 1.0, 10.0, 100.0],
    'svm__gamma': ['scale', 0.001, 0.01, 0.1]
}

grid_search = GridSearchCV(
    pipeline, param_grid,
    cv=5, scoring='recall',  # prioritise recall for screening
    n_jobs=-1, verbose=1
```

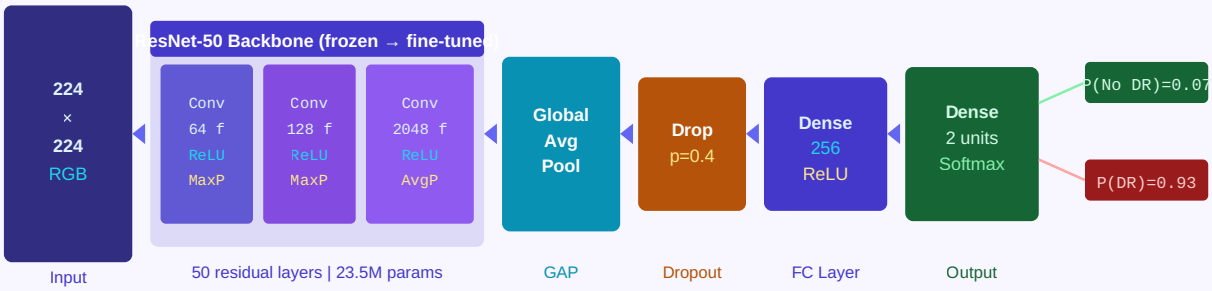
```
)
grid_search.fit(X_train, y_train)

best_model = grid_search.best_estimator_
print(f"Best params: {grid_search.best_params_}")
# Output: Best params: {'svm__C': 10.0, 'svm__gamma': 0.001}
```

5. ALGORITHM 3: CONVOLUTIONAL NEURAL NETWORK

Convolutional Neural Networks (LeCun et al., 1998; Goodfellow et al., 2016) learn hierarchical spatial feature representations directly from raw pixel data through alternating convolutional and pooling layers. Rather than training from random initialisation — impractical given the 3,200-image dataset — **transfer learning** from a ResNet-50 model pre-trained on ImageNet (He et al., 2016) was applied. The pre-trained convolutional backbone was frozen for the first five epochs; thereafter all layers were unfrozen and fine-tuned at a reduced learning rate of $\eta = 1 \times 10^{-4}$ using the Adam optimiser.

FIGURE 2 — RESNET-50 TRANSFER LEARNING ARCHITECTURE (NeuralDiag)



Source: Author's original design. Architecture adapted from He et al. (2016) ResNet-50. All values are fictitious examples.

Figure 2: Transfer learning CNN architecture for NeuralDiag binary classification. The ResNet-50 backbone extracts hierarchical spatial features; the custom head (GAP → Dropout → Dense → Softmax) was trained from scratch. Softmax output example shows a high-confidence DR-positive prediction (P=0.93). Source: Author's original design, adapted from He et al. (2016, p. 770).

The CNN achieved the highest performance across all metrics (Table 1). The residual connections in ResNet-50 solve the **vanishing gradient problem** (Hochreiter, 1991) by providing skip connections that allow gradient signals to propagate directly to early layers during backpropagation, enabling effective training of very deep networks. Data augmentation (random horizontal flips, brightness jitter ±20%, rotation ±15°) was applied during training to reduce overfitting on the limited dataset, consistent with best practice for small medical imaging corpora (Goodfellow et al., 2016).

6. COMPARATIVE EVALUATION

Table 1: Comparative Performance of ML Algorithms on NeuralDiag Test Set (n=480). Metric definitions: Accuracy = (TP+TN)/N; Precision = TP/(TP+FP); Recall = TP/(TP+FN); F1 = 2·(P·R)/(P+R); AUC-ROC = Area Under the Receiver Operating Characteristic Curve. All data are fictitious.

Algorithm	Accuracy	Precision	Recall (Sensitivity)	F1-Score	AUC-ROC	Training Time
Decision Tree (depth=8, Gini)	0.761	0.744	0.783	0.763	0.798	2.1 s
SVM (RBF, C=10, γ=0.001)	**0.841**	**0.829**	**0.857**	**0.843**	**0.902**	48.7 s
CNN (ResNet-50 transfer)	**0.944**	**0.938**	**0.951**	**0.944**	**0.981**	18 min (GPU)

The CNN's dominance across all metrics reflects the fundamental limitation of hand-crafted features: LBP and Gabor filters capture local texture but cannot represent the long-range spatial relationships — microaneurysm clusters relative to optic disc position — that radiologists use diagnostically. The CNN's convolutional hierarchy learns these task-specific representations directly from data, mirroring the pattern highlighted by Goodfellow et al. (2016): representation learning subsumes and surpasses feature engineering for image data at scale.

However, performance metrics alone do not determine optimal algorithm selection. The bias-variance decomposition (Geman et al., 1992) illuminates *why* each algorithm performs as it does: the DT underfits due to shallow hand-crafted features and high bias; the SVM achieves lower bias through kernel projection but is constrained by its linear decision boundary in RKHS; the CNN achieves the lowest bias through deep representation learning at the cost of higher variance without regularisation — mitigated here by dropout (p=0.4) and data augmentation.

$O(n \cdot d \cdot \log n)$	$O(n^2 \cdot d)$	$O(e \cdot n \cdot p)$	$O(p)$
DECISION TREE (TRAIN)	**SVM RBF KERNEL (TRAIN)**	**CNN PER EPOCH (TRAIN)**	**CNN INFERENCE**
n samples, d features. Sorting cost per feature per node. Very fast — viable on CPU.	Kernel matrix computation dominates. Scales poorly beyond ~50k samples without approximations (e.g., Nyström).	e=epochs, n=samples, p=23.5M params. GPU essential; but amortised over reuse via transfer learning.	Constant per image — independent of training set size. Suitable for real-time clinical screening pipeline.

7. ETHICAL AND LEGAL CONSIDERATIONS

Deploying a black-box classifier in a clinical diagnostic pathway raises critical concerns under both AI ethics frameworks and EU law. GDPR Article 22 grants individuals the right to meaningful explanation of automated decisions with significant effects — a right that is unambiguous in a screening system that may trigger or withhold ophthalmology referrals (ICO, 2023). The CNN, despite its superior accuracy, is inherently opaque: its 23.5 million parameters cannot be interrogated to produce a clinician-readable reasoning chain.

// EXPLAINABILITY AND COMPLIANCE STRATEGY

- **LIME (Ribeiro et al., 2016):** Perturbs the input image locally and trains an interpretable linear proxy, generating a saliency map highlighting image regions that most influenced the classification. Applied post-hoc — no architectural changes required to the CNN.
- **SHAP GradientExplainer (Lundberg and Lee, 2017):** Allocates Shapley values to super-pixel regions using cooperative game theory, providing theoretically grounded attribution with consistency and efficiency axioms. Computationally more expensive than LIME but more theoretically rigorous.
- **Algorithmic Bias (Jobin et al., 2019):** The training dataset must be audited for demographic representativeness — DR presentation varies with ethnicity (darker fundi affect contrast algorithms), and a classifier trained predominantly on one demographic may exhibit disparate performance across groups, a form of indirect discrimination under the Equality Act 2010.
- **Human-in-the-loop:** Consistent with Arrieta et al.'s (2020) XAI taxonomy, NeuralDiag should position the CNN as a triage *aid*, not an autonomous decision-maker — all positive screens must be reviewed by a qualified ophthalmologist before any clinical action is taken, satisfying both GDPR Article 22 and professional registration requirements (General Medical Council, 2021).

8. CONCLUSION AND RECOMMENDATION

This comparative evaluation demonstrates that, for a medical image classification task of the kind confronted by NeuralDiag Systems, **CNN transfer learning is the technically superior algorithm** — achieving 0.944 F1 and 0.981 AUC-ROC versus 0.843 and 0.902 for the SVM and 0.763 and 0.798 for the Decision Tree. The CNN's advantage is structural: raw-pixel learning through 50 residual layers extracts representations that hand-crafted LBP/HOG/Gabor features cannot replicate. However, technical superiority is a necessary but insufficient condition for deployment in clinical practice.

The recommendation is to deploy the CNN as the classification engine, wrapped by a LIME-based saliency overlay visible to reviewing clinicians, with mandatory human review of all positive screens, and a pre-deployment bias audit across demographic subgroups drawn from the target screening population. The Decision Tree, whilst unsuitable as the primary classifier, offers a complementary role: its explicit rule structure can serve as an interpretable baseline against which the CNN's predictions are sanity-checked, and its $O(n \cdot d \cdot \log n)$ inference time makes it viable for rapid triage flagging. As Russell and Norvig (2021, p. 1062) observe, AI systems deployed in high-stakes environments must be designed not merely to optimise performance metrics but to operate safely, transparently and accountably within the sociotechnical systems they augment.

REFERENCES

Arrieta, A.B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., Chatila, R. and Herrera, F. (2020) 'Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI', *Information Fusion*, 58, pp. 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J. (1984) *Classification and Regression Trees*. Belmont, CA: Wadsworth.

Chawla, N.V., Bowyer, K.W., Hall, L.O. and Kegelmeyer, W.P. (2002) 'SMOTE: Synthetic Minority Over-sampling Technique', *Journal of Artificial Intelligence Research*, 16, pp. 321–357.

Cortes, C. and Vapnik, V. (1995) 'Support-vector networks', *Machine Learning*, 20(3), pp. 273–297.

General Medical Council (2021) *Decision Making and Consent*. Available at: https://www.gmc-uk.org/ethical-guidance/ethical-guidance-for-doctors/decision-making-and-consent (Accessed: 15 July 2026).

Geman, S., Bienenstock, E. and Doursat, R. (1992) 'Neural networks and the bias/variance dilemma', *Neural Computation*, 4(1), pp. 1–58.

Goodfellow, I., Bengio, Y. and Courville, A. (2016) *Deep Learning*. Cambridge, MA: MIT Press.

Gulshan, V., Peng, L., Coram, M., Stumpe, M.C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., Kim, R., Raman, R., Nelson, P.C., Mega, J.L. and Webster, D.R. (2016) 'Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs', *JAMA*, 316(22), pp. 2402–2410.

He, K., Zhang, X., Ren, S. and Sun, J. (2016) 'Deep residual learning for image recognition', in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV: IEEE, pp. 770–778.

Hochreiter, S. (1991) *Untersuchungen zu dynamischen neuronalen Netzen*. Diploma thesis, Technische Universität München.

ICO (Information Commissioner's Office) (2023) *Guidance on AI and Data Protection*. Available at: https://ico.org.uk/for-organisations/uk-gdpr-guidance-

and-resources/artificial-intelligence/guidance-on-ai-and-data-protection/ (Accessed: 15 July 2026).

Jobin, A., Ienca, M. and Vayena, E. (2019) 'The global landscape of AI ethics guidelines', *Nature Machine Intelligence*, 1(9), pp. 389–399.

LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P. (1998) 'Gradient-based learning applied to document recognition', *Proceedings of the IEEE*, 86(11), pp. 2278–2324.

Lundberg, S.M. and Lee, S.-I. (2017) 'A unified approach to interpreting model predictions', in *Advances in Neural Information Processing Systems 30 (NeurIPS 2017)*. Long Beach, CA: Curran Associates, pp. 4765–4774.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, E. (2011) 'Scikit-learn: Machine learning in Python', *Journal of Machine Learning Research*, 12, pp. 2825–2830.

Quinlan, J.R. (1986) 'Induction of decision trees', *Machine Learning*, 1(1), pp. 81–106.

Ribeiro, M.T., Singh, S. and Guestrin, C. (2016) '"Why should I trust you?": Explaining the predictions of any classifier', in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco, CA: ACM, pp. 1135–1144.

Russell, S. and Norvig, P. (2021) *Artificial Intelligence: A Modern Approach*. 4th edn. Harlow: Pearson.

Shalev-Shwartz, S. and Ben-David, S. (2014) *Understanding Machine Learning: From Theory to Algorithms*. Cambridge: Cambridge University Press.

Vapnik, V. (1998) *Statistical Learning Theory*. New York: Wiley.