A Survey on Rice Grain Classification from Traditional Methods to Deep Learning Approaches

M. Niranjana and F. Kurus Malai Selvi
Government College for Women (A) (Affiliated to Bharathidasan University, Tiruchirappalli) Kumbakonam, Tamil Nadu, India

Keywords: Raw Rice, Classification, Machine Learning Deep Learning, Ensemble Model.

Abstract: Challenges persist in developing a suitable method to distinguish cultivated quality rice seeds, which can be estimated based on their characteristics. To avoid rice grain varieties from getting incorrectly labelled, the quality and types of rice grains must be identified. In this paper, classification rice grains are analysed and study is done on different types of algorithms for every stage. Generally, visual observations are made with specialists using specific devices measuring various properties. The resultant data are fed into different stages using various algorithms which are discussed in detail. This study reviews machine learning techniques to differentiate between rice seeds using different types of algorithms. Every stage is analysed under different objectives and important conclusions that gives extensions to the next stage of the research.

1 INTRODUCTION

Rice is a vital staple crop, serving as a primary food source for over half of the world's population (Heiser et al., 1993). Following wheat and corn, it ranks as the third most cultivated and consumed cereal globally.

In many regions, particularly in Asia and parts of Africa, rice represents the principal source of both dietary protein and caloric energy (Murshed, Muntasir, and Muntaha Masud Tanha., 2021).

The economic significance of rice is underscored by the fact that one-third of global rice consumption is facilitated through international trade, with over 60% of the world's population residing in Asia-where rice is a dietary staple (Chatnuntawech I et al.,). Beyond its economic value-generating employment and foreign revenue-rice also provides numerous nutritional and health benefits. It is a valuable source of vitamin B1 and contributes positively to blood sugar regulation, digestive health, and aging prevention. Furthermore, rice is extensively used in various industrial applications due to its high starch content (Chatnuntawech I et al.,).

Rice is the principal staple food crop in South India, playing a critical role in regional food security. Various rice varieties are cultivated across different parts of the country to meet the nutritional demands of the growing population (Khanam, Rubina, et al., 2020).

Farmers consistently encounter irrecoverable losses due to multiple reasons such as climate change, drought, and seed quality issues. Currently, the Seed Testing Laboratories (STL) are responsible for certifying seed quality, with trained technicians conducting purity tests manually (Rajalakshmi, Ratnavel, et al., 2024).

However, seed classification lacks uniformity across various laboratories due to factors such as technician fatigue, eye strain, and individual circumstances (Hilton, Susan., 2018). Therefore, the automation of rice seed variety identification is essential for guaranteeing the quality and germination potential of rice crops.

As highlighted in (Patrício, Diego Inácio, and Rafael Rieder, 2018), the classification of rice grains is essential due to the wide variety of rice types available in the market. Manual classification, however, is labor-intensive and time-consuming, often leading to inconsistencies. To address this challenge, the development of intelligent automated systems is necessary to enhance efficiency and accuracy in rice grain classification.

The result of these features is analysed using one of the machine learning techniques. This study identifies classified rice images under the following steps, they are,

a. Images are captured using camera.
b. Captured images are undergone pre-processing techniques to sharpen the image quality.

Niranjana, M. and Selvi, F. K. M.
A Survey on Rice Grain Classification from Traditional Methods to Deep Learning Approaches.
DOI: 10.5220/0013912200004919

In *Proceedings of the 1st International Conference on Research and Development in Information, Communication, and Computing Technologies (ICRDICCT'25 2025) - Volume 4*, pages 311-320
ISBN: 978-989-758-777-1

c. Feature extraction is performed on the pre-processed image.
d. Extracted features are fed to the selected machine learning algorithms for classification.

This system should be capable of automatically identifying and categorizing individual rice grains. The primary process involves collecting a dataset and extracting various parameters of individual rice grains, such as major and minor axis, eccentricity and length, and also the breadth.

2 LITERATURE REVIEW

The input image is ultimately utilized as training data, where each grain of rice within the image is mapped to its corresponding class for classification purposes (Kiratiratanapruk, Kantip, et al., 2020)

Some traditional rice varieties of India cultivated in different regions are Basmati, Joha, Jyothi, Navara (Special varieties), Ponni, Pusa, Sona Masuri, Jaya (Intermediate-varieties) Kalajiri (aromatic), Boli Palakkad Matta etc.

Some varieties of colored rice are grown in the country include Himalayan red rice; Matta rice, Kattamodon, Kairali, Jyothy, Bhadra, and Asha in Kerala; Rakthashali in Kerala; and Red Kavuni, Kaivara Samba, Mappillai Samba, Kuruvi kar and Poongar rice in Tamil Nadu. The classification of such diverse rice types plays a critical role in agricultural research and quality assessment. However, manual classification is laborious, time-consuming, and prone to human error. Therefore, the adoption of intelligent automated systems has become essential to streamline the classification process.

Any rice sample analyzed by an automated system typically undergoes a structured sequence of steps, including classification, segregation, evaluation, categorization, and grading. The primary contribution of this study lies in categorizing existing algorithms and techniques into five major approaches: geometric, statistical, supervised learning, unsupervised learning, and deep learning.

Deep learning techniques, in particular, have produced promising findings and have sparked interest in future. Also, these techniques can able to separate rice grain into different classes efficiently (Rathna Priya, T. S., et al., 2017)

To guarantee a good yield and quality, rice types must be accurately identified. Grain attributes such as color, shape, taste, aroma, cooking characteristics, and head rice recovery are analyzed, together with morphological features and visual inspection, as traditional methods of rice variety identification) Alfred, Rayner, et al., 2021)

Some key research questions that rose during the study of the research topic. They are,

a) What methods and algorithms have been previously suggested for categorising rice?
b) What are all the steps taken to identify the most suitable method among the different types of algorithms?
c) Differentiation between manual rice grading and automated rice grading approach.
d) What could be the best fit approach which can be extended for further research?

A detailed survey of rice classification techniques is presented in separate subsections. Section III provides a comprehensive overview of various learning approaches used for classifying different rice varieties, offering insights that may guide future research in this domain.

Numerous research papers related to rice grain classification were analysed as part of this study. The collected articles are systematically categorized based on the underlying methodologies, including geometric, statistical, machine learning, and deep learning approaches. Additionally, the study reviews a range of algorithms and techniques employed for rice grain detection and categorization.

Although traditional feature-based recognition methods have yielded promising results, they often rely on highly specific features. These features may fail to capture the intrinsic characteristics of rice grains, leading to limitations in classification performance.

Methods of classifying collected research articles have been explained in detail under stages like screening, Eligibility analysis, Extraction of data.

Evaluating the quality of rice grains is essential to meet consumer expectations, and grain quality is primarily determined by geometric characteristics. In local industries, mechanical classification methods are commonly employed to rank food grains based on geometric parameters. However, image processing techniques offer a more versatile and efficient alternative, enabling the extraction of various shape- and geometry-based features for rice grain classification.

Zhao et al. extracted eleven geometric properties from binary images of rice kernels, including perimeter, area, circularity, equivalent diameter, number of contour points, major axis length, minor axis length, rectangle elongation, maximum inscribed circle, and minimum enclosing circle. In addition to geometric features, texture features were extracted from grayscale images such as mean, variance, smoothness, consistency, entropy, and seven statistically invariant moments.

In a related study, (Barbedo, J.G.A., 2016) utilized textural features for corn image classification, including energy, contrast, homogeneity, correlation, and Local Binary Pattern (LBP)-based Gary Level Co-occurrence Matrix (GLCM). These feature extraction techniques can be broadly categorized into geometry-based, statistics-based, and learning-based methods. Learning-based approaches include both unsupervised techniques-such as k-means clustering for grouping unlabelled data-and supervised models like neural networks, support vector machines (SVM), and, more recently, deep learning architectures.

Among all the approaches, the supervised approach contributed maximum share. Better performance of supervised approaches can be attributed to use of handcrafted spatial features along with different classifiers.

Based on the knowledge gained, further learning approaches of machine learning were analysed in section A.

3 MACHINE LEARNING APPROACHES

Machine learning approach can be broadly classified into supervised, unsupervised, and deep learning.

3.1 Unsupervised Learning

Large datasets are not necessary for the classifier to be trained using unsupervised learning approaches. Table 1 tabulates the different types of These methods use clustering to divide the data into classes according to how similar they are. A clustering-based method for classifying rice is demonstrated in (T. Bera et al., 2019).

The authors captured two images of a rice sample consisting of eight different rice varieties. After that, the photos underwent pre-processing. using edge detection methods, thresholding, and the elimination of noise and lens distortion. One such a PCA-based approach to categorizing various Basmati rice varieties was introduced in (T. Bera et al., 2019) which, rather than using dendrograms, used clustering based on K-Nearest Neighbours (KNN) which is a basic machine learning algorithm that considers the similarity between the new case/information and available cases and assigns the new case to the class that is most similar to the available classes.

The rice image was pre-processed using KNN clustering, noise reduction, and smoothing. In this paper, has experimented with six different types of rice seeds. After segmenting the coloured rice image, binarization was performed. Morphological features like order were used to improve and make sense of the input image after the aforementioned operations were completed. Area, major axis length, minor axis length, eccentricity, and perimeter were among the parameters used to extract the image. The various rice grain varieties were then clustered using KNN as a classifier. Along similar lines, the authors devised a rice quality classification system that also used KNN clustering. Techniques with extracted features of geometric, texture and shape feature extraction techniques.

Table 1: Feature based analysis on unsupervised and Geometric features.

Author	Algorithm/Techniques	Extracted Features
Mahajan, S(2015)	edge detection techniques	Geometric Features and Texture features
Barbedo, J.G.A., (2016)	GLCM and LBP	Geometric Features and Texture features
J. P. Shah(2016)	edge detection techniques	Shape features
T. Bera(2016)	PCA based algorithm + KNN	Shape features

3.2 Supervised Learning

It is a type of Artificial Intelligence where machines are being trained in training data, and based on that data machines predicted the outcome. The mentioned data seems to be legitimate data that is currently tagged with the correct outcome.

Supervised Learning In supervised learning, the training data is given to the machines in such a way that the supervisor works (Supervises) the machines for Correct output. Morphological features such as area of the seed, seed boundary, bounding box around seed, width, length of major and minor axes, thinness ratio, aspect ratio, rectangular aspect ratio, equivalent diameter, filled area, area under major axis of ellipse, convex area, solidity, and extent were systematic extracted rice image. Furthermore, different color features were also extracted including, red, green, blue colors bands hue, saturation, intensity, and standard deviation of hue.

3.3 Deep Learning Approaches

Deep learning refers to ANNs with more layers that can make use of and learn from complex datasets such as images, audio, and text. It is constructed on the basis of the early designs and models of neural networks.

During the 1980s and 1990s, more sophisticated neural network architectures were introduced, including Multi-Layer Perceptron's (MLPs) and Radial Basis Function Networks (RBFNs). However, it was not until the early 2000s enabled by the availability of high-performance computational resources and large-scale datasets-that deep learning gained significant momentum and practical success (Naser, Samy Abu., 2018).

Deep learning models have the ability to learn abstract level representations across different types of input data such as text, image, audio and perception signals. These models have already demonstrated promising potential in rice grain classification. The deep learning algorithms and the common extracted feature methods for different studies are summarized in Table2.

Among these, one of them and perhaps the most famous and effective architecture, is the Convolutional Neural Network (CNN), that will be discussed next in detail.

Convolutional neural network (CNN): It is a type of deep learning neural network specialized for structured arrays of data, such as images. Some of the applications are:

- The effectiveness of CNNs can be shown as CNNs are widely applied in many vision applications and its architecture became a state of-the-art in many visual applications, like image classification etc.
- CNNs have also been successful in various natural language processing tasks, especially text classification (Abu Naser, S. S. and M. J. Al Shobaki, 2016).
- A CNN is a complex "deep learning neural network" specialized to process and identify images. It is particularly well at recognizing patterns in an input image, e.g., circles, lines, faces, eyes, and gradients. Abu Naser, S. S. (2012).
- Because of this, CNN is extremely efficient in the domains of computer vision.
- CNNs are feed-forward neural networks composed of multiple types of layers including convolutional layers and others; some models have as many as 20 or 30 layers.
- CNNs take advantage of the power of convolutional layers, which can learn to detect increasingly complicated shapes when stacked on top of one another.

Scratch Net has a simple architecture to make understanding the concept clearer and the applications added to CNN gave an overall idea of its creation.

1) Convolutional layer is the fundamental layer that is used to separate the various features from the input image, where convolution is performed between the input image and a channel of certain size Maxim. The outcome is called the Element map which provides us information regarding the photo such as the corners and edges. Then this part map is used in various layers for them to learn a few different parts of the input image.

The convolution layer in CNN passes the result to the next layer. Abu Naser, S. S. (2012).

From the papers J. Abu Naser, S. S. (2008). some key points were founded which are being cascaded as under.

2) Pooling layer decrease the information volume and boundary count, which can help to prevent over fitting and further develop network execution. Further, pooling is separated in two classes:

a) Max Pooling- The maximum value within the image covered by the kernel is provided by the Max Pooling, and

b) Average Pooling- The average value within the kernel's hidden image is provided by the Average Pooling function.

3) Fully Connected layer a layer that is entirely connected needs to be flattened. Before being fed into the neural network, the entire pooling feature map matrix is transformed into a single column.

function such as sigmoid. Convolutional layers apply filters or kernels to the input data, resulting in the generation of feature Devi, T. G., Neelamegam, P., & Sudha, S., (2017).

Thus, creation of a model is obtained by combination of the obtained feature map matrix using attributes of completely linked layers. The output is then classified using an activation.

Table 2: Presents A Summary of Deep Learning Algorithms.

Author	Algorithm/Techniques	Extracted Features
T. Bera(2019)	PCA based algorithm + KNN + Neural networks	Shape features
Devi, T(2017)	Machine vision algorithm	Physical and chemical features
Patel, N(2017)	Neural networks & Support Vector Machine	Shape features
Wah, T(2018)	NB tree & SMO classifiers	Chalkiness and whiteness of grains along with other morphological features
Kuchekar, N. A(2018)	Random Forest classifier, decision tree classifier	Shape and Geometric features
Son, N(2019)	Multiclass SVM	Shape descriptors, color descriptors
Han, Xu, et al(2021)	BPNN	18 color features and 21 texture features

A study in (Kuchekar, N. A., & Yerigeri, V. V., 2018). proposed new features to classify nine varieties of rice grains, getting an accuracy of 95.78% with the NB Tree and Sequential Minimal Optimization (SMO) classifier. Chalkiness and whiteness markers were extracted as features, along with important morphological features like major and minor axes, area, and perimeter.

In a separate study (Son, N. H., & Thai-Nghe, N. 2019), twenty features were extracted using a Random Forest classifier, and its performance was evaluated using standard metrics such as accuracy, precision, recall, and F1-score. The results were compared against those obtained using decision tree classifiers. Experimental findings revealed that the Random Forest classifier significantly outperformed the decision tree model, achieving a highly promising classification accuracy of 99.85%.

In work (Han, Xu, et al., 2021), a set of shape descriptors and color descriptors were extracted from every rice grain image and three types of classification was done using Multi-class Support Vector Machine which are basmathi, ponni and brown rice. The classification accuracy obtained is 92.22%.

To evaluate and categorize rice grains, their physical and chemical features were extracted using the Machine vision algorithm which averages the values of the extracted features which were considered for grading and quality analysis of rice grains Patel, N., Jayswal, H., & Thakkar, A., (2017).

In Wah, T. N., San, P. E., & Hlaing, T. (2018). the extracted shape features of rice is evaluated based on Neural Networks (NN) and Support Vector Machine (SVM) algorithms and the results infer that SVM based classification outperforms the neural networks.

Figure 1 shows proposed idea of using CNN model for classifying Kauvuni Rice. The idea has been taken by adding three convolutional layers and three pooling layers along with fully connected layers which are going to be implemented to classify the rice.

Scratched Net To extend the research, the review has been done in leading pre-trained models which pave the way to combine the existing architecture with some leading pre-trained models to propose the new scratched Net. The ResNet is a variation of CNN that builds upon the concepts introduced in Inception Networks and Residual Networks (ResNets). It exemplifies the idea of cardinality for performance and introduces a divide, transform and merge block.

Various ResNet variants have been proposed to balance performance and computational cost on a specific task and dataset (Szegedy, Christian, et al., 2017).

Pre-Trained model: These types of models are neural networks are used to train on large datasets and used for specific tasks (Han, Xu, et al., 2021). These models capture convoluted patterns and features, which is used to classify the selected features effectively.

In this paper popular models like VGG, ResNet, and Inception have set benchmarks in the field.

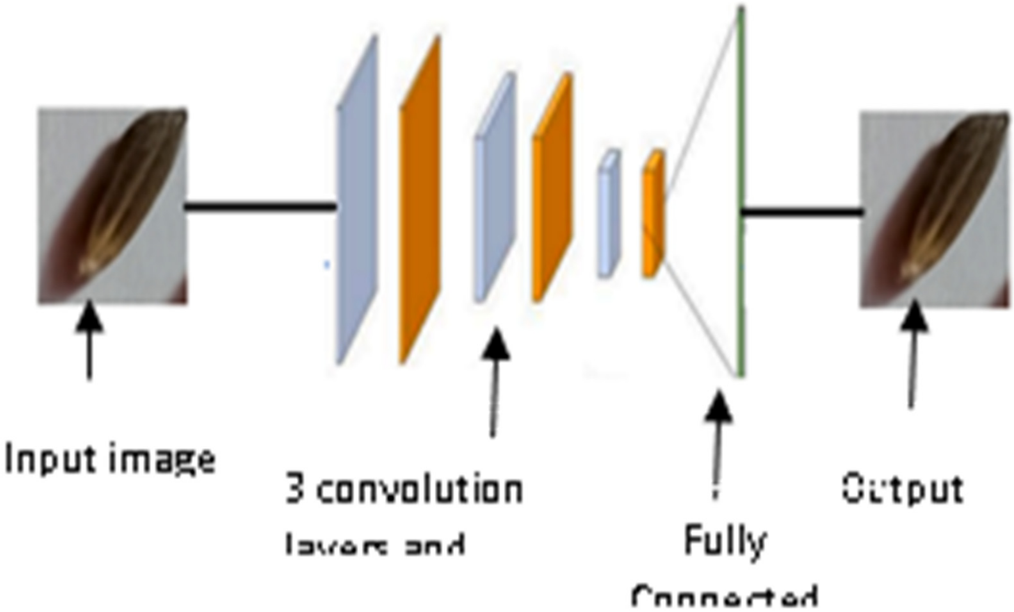

Figure 1: Proposed of CNN model.

Table 3: Top Pre-Trained Models with Scratch Net models.

Models	Variants	Key Features
ResNet (Residual Networks)	ResNet-50, ResNet-101, ResNet-152.	Deep architectures (up to 152 layers). Residual blocks to allow gradients to flow through shortcut connections.
Inception	Inception v3, Inception v4, Inception-ResNet.	Inception modules with convolutional filters of multiple sizes.
VGG (Visual Geometry Group)	VGG-16, VGG-19.	Deep networks with 16 or 19 layers.
EfficientNet	EfficientNet-B0 to EfficientNet-B7.	Compound scaling method to scale depth, width, and resolution. Efficient and accurate.
DenseNet (Dense Convolutional Network)	DenseNet-121, DenseNet-169, DenseNet-201.	Dense connections to improve gradient flow and feature reuse. Reduces the number of parameters compared to traditional convolutional networks.
MobileNet	MobileNetV1, MobileNetV2, MobileNetV3.	Lightweight architecture optimized for mobile devices. Depthwise separable convolutions.
NASNet (Neural Architecture Search Network)	NASNet-A, NASNet-B, NASNet-C.	Automatically designed architectures using reinforcement learning. High accuracy with efficient performance.
Xception (Extreme Inception)	--	Fully convolutional architecture.
AlexNet	------	Simple architecture with 8 layers. ReLU activation functions and dropout regularization.
Vision Transformers (ViT)	------	Transformer encoder architecture. Scales well with large datasets and computational resources.

Table 3 shows the Scratch Net models which are the top pre-trained models which is used nowadays to classify the features efficiently. Among the several derived nets ResNet is discussed in detail.

In (Szegedy, Christian, et al., 2017) ResNet has been discussed. This paper suggests that even in very deep topologies, the network can more readily collect and propagate gradients through the network by employing residual layers.

ResNet employs skip connections or shortcut connections or identical mappings: these connections facilitate the flow of gradients directly from the end layers to the earlier layers -- therefore bypassing the intermediate layers.

It makes it easier for the gradients to move backwards since they are not getting smaller and smaller as they back propagate through the network. Res Net also uses skip connections which allow the network to be much deeper than previous architectures. About the Author Res Net models are built successfully with depths of 50, 101, or even 152 layers.

One significant development in the field of deep learning has been the capacity to train and efficiently optimize such deep networks. Res Net reduces the vanishing gradient issue and makes training very deep neural networks easier by utilizing residual layers and skip connections. Image classification, object detection, and image segmentation are just a few of the computer vision tasks where this has improved performance and accuracy (Shafiq et al., 2022).

The above analysis paved a way to create own model combining one or two of the above said models which may produce effective result. Also, these types of hybrid models are otherwise known to be Ensemble model. This Ensemble model aims to improvise the prediction accuracy, and ultimately reduces generalization error.

Ensemble model: A machine learning method called ensemble learning blends w models to increase prediction accuracy:

To provide more accurate predictions, ensemble learning combines several learners, such as neural networks or regression models. One student is believed to be less accurate than a group of students. Ensemble learning can be used to address problems such as excessive variance, under fitting, and over fitting. Classification, regression, and clustering are just a few of the machine learning tasks that can benefit from ensemble learning (Yadav, Pravin Singh, et al., 2025)

Additionally, it can be applied to short-term forecasting, landslide assessment, and incursion detection. Voting is used in ensemble learning models to decide the final result. Ensemble learning models have shown significant promise across various domains, including short-term forecasting, landslide assessment, and incursion detection. These models leverage the principle of voting among multiple classifiers to determine the final outcome, thereby improving prediction reliability and robustness.In (Hasan et al., 2023), an enhanced ensemble approach—Ensemble Model of the Stowing and Helping Classifier (EMBBC)-was proposed for the prediction of code smells. This model integrates decision-making and data-balancing strategies to improve classification performance. The experimental analysis was conducted using four publicly available code smell datasets: Blob Class, Information Class, Long Boundary Rundown, and Switch Statement. To address class imbalance, the datasets were preprocessed using the Synthetic Minority Over-sampling Technique (SMOTE)Similarly, the study in (Tkatek, S et al., 2023) introduced an ensemble machine learning framework combining K-Nearest Neighbors (KNN), Random Forest (RF), and Kernel Ridge Regression (KRR). The model's performance was evaluated using four standard evaluation metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R^2 score. The results demonstrated that the proposed KRR-based ensemble method achieved superior accuracy in yield prediction when compared with other conventional machine learning models.

Work in (Long, C et al., 2024) tested ML models (eXtreme gradient boosting (XG Boost), K-Nearest Neighbor (KNN), Linear support vector machine (SVM), Naive Bayes (NB) classifier, decision tree (DT) regressors, and random forest (RF) regression) to anticipate the potato crops that are both of excellent and yield, and they utilized measurements like MAE, MSE, RS, and RMSE for assessment.

Table 4 shows some Ensemble models that have been analysed during this study. This exploration analyses around different combinations of machine learning algorithms as Ensemble to highlight the best choice among the different procedures to find out the best possible Ensemble solution for further research. (Hasan, M et al., 2023)

Chalkiness is a crucial determinant of rice quality, assessed through three key indicators: chalkiness size, chalky rice rate, and chalkiness degree. The chalky rice rate indicates the proportion of chalky grains, while chalkiness size measures the percentage of chalky area within each kernel. Chalkiness degree is the product of these two factors. In a recent agricultural experiment in China, human visual inspection is used to evaluate chalkiness. While this method is accurate for determining the chalky rice rate, it can lead to significant errors in measuring chalkiness size and requires substantial manpower. Adopting more efficient evaluation methods could improve rice quality control and benefit both producers and consumers. Considering the above factors review has been made as under.

The authors in (Tkatek, set al., 2024) examined the chalkiness, fine starch structure, and physiochemical properties of rice, and correlated and additional studies were conducted under varying nighttime temperatures throughout the early grain-filling stage. Higher chalky grain rate (CGR) and chalkiness degree (CD) were induced by medium temperature (MT) and low (LNT) and high nighttime temperatures (HNT) when compared to MT.
LNT mainly improved the chalkiness by increasing short branch chains of amylopectin, increasing the degree of branching and the ratio of small starch granules, while reducing Long Branch chains of amylopectin and amylose branches. It changed the pasting properties, for example, the peak viscosity and final viscosity were increased.

In (Long, C et al., 2024), the distance between the chalky areas and the minimum enclosing rectangle of the resulting rice kernel after separating the residual embryo, back white, core white, mid white through the Support Vector Machine (SVM) classifier in a very careful manner was studied. This innovative approach not only enables the clear differentiation of embryos from chalky regions but also facilitates the complete removal of residual embryos while achieving remarkable accuracy in pinpointing chalkiness locations. The results of this research lay a robust theoretical and practical foundation for the advanced application of computer vision technology in the detection of chalkiness, promising to revolutionize quality control in rice production.

To conclude, this paper attempts to review the variety of algorithms to imply ensemble deep learning. The contributions of this paper are highlighted as the following.

First, analysis is made on unsupervised learning Secondly introduced the basic concepts of supervised learning and their advantages are discussed with their advantageous clearly. Thirdly deep learning approaches are discussed along with basic architecture of CNN and the advantages.

Table 4: Ensemble models.

Author	Ensemble Model Name	Combinations
Yadav, Pravin Singh, et al(2025)	EMBBC MODEL	Bagging and Boosting Classifiers
Hasan, M(2023)	K-Nearest Neighbor, Random Forest, Ridge Regression (KRR)	Mean absolute error, Mean Square error (MSE), Root mean square Error (RMSE), and R2
Tkatek, S(2023)	XGBoost	K-Nearest Neighbor (KNN), Linear support vector machine (SVM), Naive Bayes (NB) classifier, Decision tree (DT) regressors, and Random forest (RF) regression

Moreover, this paper discusses the different strategies of ensemble deep learning models. Finally, comprehensive review has made to elaborate the proposed research work.

4 CONCLUSIONS

This paper reviews algorithms based on geometric algorithms, machine learning, and deep learning models and also the Ensemble model. These reviews gave concrete ideas to extend the study into the next level. This work has been ignited to classify the results with the suitable algorithms are analysed and identified images of rice grains are taken for further study. This study looked at different ways to classify things, including detection techniques, Gray-Level Co-Occurrence Matrix (GLCM) and Local Binary Pattern (LBP), PCA based algorithm combined with KNN. It also analyses at different ways to choose features, including geometric, texture and shape features.

This study further extends its analysis by evaluating several machines learning models, including combinations such as Principal Component Analysis (PCA) with K-Nearest Neighbors (KNN) and Neural Networks, as well as hybrid approaches like Neural Networks with Support Vector Machines (SVM), Naïve Bayes Tree (NBTree) with Sequential

Minimal Optimization (SMO), Random Forest, Decision Tree, Multiclass SVM, and Backpropagation Neural Networks (BPNN).

Furthermore, we also investigate the performance of both pre-trained and from-scratch deep learning models, including state-of-the-art models like ResNet, Inception, VGG, EfficientNet, DenseNet, MobileNet, NASNet, Xception, AlexNet, and ViTs.

The findings from this research have practical implications for software engineers and researchers, particularly in understanding how to mitigate the adverse effects of code smells on software quality.

Moreover, this study emphasizes the importance of feature selection, which plays a critical role in enhancing the effectiveness of ensemble learning strategies. These insights may inspire further exploration into optimal hybrid ensemble approaches that can significantly improve classification performance, particularly in the domain of rice grain classification.

Author Contribution

Niranjana Mahalingam: Conceptualization, Methodology, Writing - Original draft;
Kurus Malai Selvi: Review & editing, Visualization, Supervision.

Declaration of Competing Interest

Competing interests, the authors declare no competing interests.

Funding Information

Funding: This work was not supported by any specific grant from funding agencies in the public, commercial,

REFERENCES

Heiser, Charles B. *Seed to civilization: the story of food.* Harvard University Press, 1990. W.-K. Chen, *Linear Networks and Systems* (Book style). Belmont, CA: Wadsworth, 1993, pp. 123–135.

Murshed, Muntasir, and Muntaha Masud Tanha. "Oil price shocks and renewable energy transition: Empirical evidence from net oil-importing South Asian economies." *Energy, Ecology and Environment* 6.3 (2021): 183-203.

Chatnuntawech I, Tantisantisom K, Khanchaitit P, Boonkoom T, Bilgic B, Chuangsuwanich E. Rice classification using spatio-spectral deep convolutional neural network. arXiv preprint 8; arXiv:1805.11491

Khanam, Rubina, et al. "Metal (loid) s (As, Hg, Se, Pb and Cd) in paddy soil: Bioavailability and potential risk to human health." *Science of the Total Environment* 699 (2020): 134330.

Rajalakshmi, Ratnavel, et al. "RiceSeedNet: Rice seed variety identification using deep neural network." *Journal of Agriculture and Food Research* 16 (2024): 101062.

Hilton, Susan. "Improving Processing Vegetable Yields Through Improved Production Practices." (2018).

Patrício, Diego Inácio, and Rafael Rieder. "Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review." *Computers and electronics in agriculture* 153 (2018): 69-81.

Abbaspour-Gilandeh, Yousef, et al. "A combined method of image processing and artificial neural network for the identification of 13 Iranian rice cultivars." *Agronomy* 10.1 (2020): 117.

Kiratiratanapruk, Kantip, et al. "Development of paddy rice seed classification process using machine learning techniques for automatic grading machine." *Journal of Sensors* 2020.1 (2020): 7041310.

Rathna Priya, T. S., et al. "Nutritional and functional properties of coloured rice varieties of South India: a review." *Journal of Ethnic Huang, Gao, et al. "Densely connected convolutional networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.Foods* 6.1 (2019): 1-11.

Alfred, Rayner, et al. "Towards paddy rice smart farming: a review on big data, machine learning, and rice production tasks." *Ieee Access* 9 (2021): 50358-50380.

Rice varieties - IRRI Rice Knowledge Bank. (n.d.). Rice Varieties - IRRI Rice Knowledge Bank. doi: http://www.knowledgebank.irri.org/step-by-step-production/pre-planting/rice-varieties.

Mahajan, S., Das, A., Sardana, H.K., 2015. Image acquisition techniques for assessment of legume quality. Trends Food Sci. Technol. 42 (2), 116–133. https://doi.org/ 10.1016/j.tifs.2015.01

Barbedo, J.G.A., 2016. A review on the main challenges in automatic plant disease identification based on visible range images. Biosyst. Eng. 144, 52–60. https://doi.org/10.1016/j.biosystemseng.2016.01.017. Zhai, Xiaohua, et al. "Scaling vision transformers." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2022.

J. P. Shah, H. B. Prajapati, and V. K. Dabhi, ''A survey on detection and classification of rice plant diseases,'' in Proc. IEEE Int. Conf. Current Trends Adv. Comput. (ICCTAC), Mar. 2016, pp. 1–8.

T. Bera, A. Das, J. Sil, and A. K. Das, ''A survey on rice plant disease identification using image processing and data mining techniques,'' in Emerging Technologies in Data Mining and Information Security. Singapore: Springer, 2019, pp. 365–376.

Naser, Samy Abu. "An agent based intelligent tutoring system for parameter passing in java programming." *Journal of Theoretical and Applied Information Technology* 4.7 (2008).

J. Abu Naser, S. S. (2008). "Developing visualization tool for teaching AI searching algorithms." Information Technology Journal, Scialert 7(2): 350-355.

Abu Naser, S. S. (2012). "A Qualitative Study of LP-ITS: Linear Programming Intelligent Tutoring System." International Journal of Computer Science & Information Technology 4(1): 209

Abu Naser, S. S. and M. J. Al Shobaki (2016). "Enhancing the use of Decision Support Systems for Re-engineering of Operations and Business-Applied Study on the Palestinian Universities." Journal of Multidisciplinary Engineering Science Studies (JMESS) 2(5): 505-512

A. Sheeba, P.S. Kumar, M. Ramamoorthy, S. Sasikala, Microscopic image analysis in breast cancer detection using ensemble deep learning architectures integrated with web of things, Biomed. Signal Process Control 79 (2023)

Devi, T. G., Neelamegam, P., & Sudha, S., (2017). Machine vision-based quality analysis of rice grains. IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI-2017), 1052-1055.

Patel, N., Jayswal, H., & Thakkar, A., (2017). Rice quality analysis based on physical attributes using image processing technique. 2nd IEEE International conference on recent trends in electronics information communication technology, 42-47.

Wah, T. N., San, P. E., & Hlaing, T. (2018). Analysis on feature extraction and classification of rice kernels for Myanmar rice using image processing techniques. International Journal of Scientific and Research Publications, 8(8), 603-606.

Kuchekar, N. A., & Yerigeri, V. V. (2018). Rice grain quality grading using digital image processing techniques. IOSR J Electronics Communication Eng, 13(3), 84-88.

Son, N. H., & Thai-Nghe, N. (2019, November). Deep learning for rice quality classification. In 2019 international conference on advanced computing and applications (ACOMP) (pp. 92-96). IEEE.

Han, Xu, et al. "Pre-trained models: Past, present and future." *AI Open* 2 (2021): 225-250.

Shafiq, Muhammad, and Zhaoquan Gu. "Deep residual learning for image recognition: A survey." *Applied Sciences* 12.18 (2022): 8972.

Szegedy, Christian, et al. "Inception-v4, inception-resnet and the impact of residual connections on learning." *Proceedings of the AAAI conference on artificial intelligence*. Vol. 31. No. 1. 2017.

Shah, Syed Rehan, et al. "Comparing inception V3, VGG 16, VGG 19, CNN, and ResNet 50: a case study on early detection of a rice disease." *Agronomy* 13.6 (2023): 1633.

Koonce, Brett, and Brett Koonce. "EfficientNet." *Convolutional neural networks with swift for Tensorflow: image recognition and dataset categorization* (2021): 109-123.

Sinha, Debjyoti, and Mohamed El-Sharkawy. "Thin mobilenet: An enhanced mobilenet architecture." *2019 IEEE 10th annual ubiquitous computing, electronics & mobile communication conference (UEMCON)*. IEEE, 2019.

Ren, Pengzhen, et al. "A comprehensive survey of neural architecture search: Challenges and solutions." *ACM Computing Surveys (CSUR)* 54.4 (2021): 1-34.

Chollet, François. "Xception: Deep learning with depthwise separable convolutions." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.

Ismail Fawaz, Hassan, et al. "Inceptiontime: Finding alexnet for time series classification." *Data Mining and Knowledge Discovery* 34.6 (2020): 1936-1962.

Zhai, Xiaohua, et al. "Scaling vision transformers." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022.

Zhi-Hua Zhou, Ensemble Methods: Foundations and Algorithms, CRC Press, 2012.

Yadav, Pravin Singh, et al. "Ensemble methods with feature selection and data balancing for improved code smells classification performance." *Engineering Applications of Artificial Intelligence* 139 (2025): 109527.

Hasan, M., M. A. Marjan, M. P. Uddin, S. Nam, Y. Kardy, S. Ma, and Y. Nam. 2023. Ensemble machine learning-based recommendation system for effective prediction of suitable agricultural crop cultivation. Frontiers in Plant Science 14:1234555. doi:10.3389/fpls.2023. 1234555

Tkatek, S., S. Amassmir, A. Belmzoukia, and J. Abouchabaka. 2023. Predictive fertilization models for potato crops using machine learning techniques in Moroccan gharb region. International Journal of Electrical and Computer Engineering (IJECE) 13 (5):5942. doi:10. 11591/ijece. v13i5.pp5942-5950.

Long, C.; Du, Y.; Zeng, M.; Deng, X.; Zhang, Z.; Liu, D.; Zeng, Y. Relationship between Chalkiness and the Structural and Physicochemical Properties of Rice Starch at Different Nighttime Temperatures during the Early Grain-Filling Stage. Foods 2024, 13, 1516. https://doi.org/10.3390/ foods131015

Sun ChengMing, S. C., Liu Tao, L. T., Ji ChengXin, J. C., Jiang Min, J. M., Tian Ting, T. T., Guo DouDou, G. D., ... & Liang XiuMei, L. X. (2014). Evaluation and analysis the chalkiness of connected rice kernels based on image processing technology and support vector machine.