

## Framework For Building Recommender Engine to Forecast *Mangifera Indica* Foliar Pathogens

Jasmine B\*<sup>1</sup>, Shantha Mary Joshitta R<sup>2</sup>, Jenifer Jothi Mary A<sup>3</sup>

<sup>1,2</sup> Jayaraj Annapackiam College for Women (Autonomous), Periyakulam, Theni Dt., Tamilnadu, India, Affiliated to Mother Teresa Women's University, Kodaikanal

<sup>3</sup> St. Joseph's College (Autonomous), Trichy-2, Tamilnadu, India, Affiliated to Bharathidasan University, Tiruchirappalli

**Abstract:** In Asia, there are many different types of horticulture farming. The fifth-most economically significant tropical fruit in the world is the mango (*Mangifera indica* L.). The mango fruit is consumed as a fresh fruit, in juices and beverages, jams, and other foods because of its distinctive flavour and nutritional content. Mangoes are grown in more than 90 countries, with Asia producing 75% of all mangoes worldwide. There are numerous different mango species, and each has an own commercial market. Yet, a variety of illnesses have a direct impact on the quantity and quality of fruit. The primary global constraint on mango output is diseases brought on by pathogenic fungus. The most significant diseases in India include stem-end rot (*Lasiodiplodia* spp. and *Neofusicoccum* spp.), powdery mildew (*Pseudoidium anacardii*), anthracnose (*Colletotrichum* spp.), mango deformation disease (*Fusarium* spp.), and grey leaf spot (*Pestalotiopsis mangiferae*). This research study's main goal was to use machine learning techniques to analyse leaves in order to forecast future leaf diseases. The framework for developing a recommender engine to predict leaf diseases is presented in this research. It gathers information from a live test site in Western Gatz, India, and builds an expert model using the proper knowledge representation approaches. A knowledge-based recommendation engine is created for mango disease forecasting and pre-treatment as a result of this research.

**Keywords:** *Mangifera indica*; Mango leaf diseases; Recommender Engine; Western Gatz; Theni District; India,

### Introduction

The mango (*Mangifera indica* L.) is the national fruit of India and is considered the king of fruits. India is the world's largest producer of mangoes. According to the ABC Mango 2020-21 crop report, India produced 21.8 million tonnes of *Mangifera Indica* in the year 2017-18, contributing 40% of the total world production [1]. *Mangifera Indica* is cultivated in an area of 2,309,000 hectares with an annual production of 12,750,000 tons [2]. It contributes about 18.1% of India's total GDP in the agricultural sector. In Tamilnadu, the districts of Dharmapuri, Vellore, Tiruvallur, Theni and Madurai are the cultivation zones of *Mangifera Indica*. According to Theni District Statistical Manual 2020-21, *Mangifera Indica* is cultivated on 2130 hectares [3]. This paper investigates mango leaf diseases in Western Gatz of India and presents a framework for predicting some mango leaf diseases in earlier days.

### Mango Leaf Diseases

Diseases of *Mangifera Indica* plants are the great obstacle to producing enough fruit to meet people's demand. There are many diseases, such as mango malformation, anthracnose, bacterial flower disease, Golmachi, Moricha disease, Shutimold, bacterial black spot, apical shoot necrosis, red rust, lichens, powdery mildew, root rot and *Ganoderma* root rot affect mango trees [4]. On the other hand, factors such as climate change and soil infertility also affect mango production. Diseases cause significant production losses and economic losses. The main reason for this decline is the difficulty for farmers to determine the diseases affecting mango trees. The only effective method to reduce losses from most mango diseases is prevention. Therefore, it is very important to identify diseases in their early stages to safeguard trees and avoid economic losses and increase crop productivity for farmers.

Therefore, it is very important that trees are tested before effective control actions are taken.

There is a great demand for any technological approach to predict the diseases that affect mango trees. This research seeks to present a framework for the recommendation system to predict mango leaf diseases in its early days.

### **Related Research Works**

S Arya et al., (2019) presented a comparative study of CNN (Convolutional Neural Network) and Alex Net for disease detection in potato and mango leaves [5]. A dataset with 4004 images was used in this study, of which 3523 were used for training and validation and 481 images were used to test the results. CNN training accuracy is 97.75%. The Alex Net architecture achieved the highest accuracy at 98.33% against CNN. Mustafa et al., (2018) proposed the M-Means Cluster algorithm for deficiencies detection in mango leaf [6]. Symptoms of various natural leaf colors were extracted as RGB components from healthy leaves. Researchers used feature extraction technique based on K-means clustering algorithm and automatic deficiencies detection. They suggested farmers take countermeasures to minimize the risk of unhealthy plant growth.

Prakash et al., (2015) proposed an algorithm based on K-Means clustering technique for the segmentation and BPNN for the classification of the mango leaf affected by the bacterium [7]. They tested four different numbers of clusters. In the first phase, all three segments of the group produced 70% accuracy, while all five segments of the group produced 100% accuracy. In the second step, bacterial leaf spot was identified with 90% accuracy. In the final step, they found that scarlet rust was identified with 40% accuracy in 3 clusters, 90% in 5 cluster segments, 80% in 7 cluster segments, and 100% in 10 cluster segments.

Kien Tray et al., (2019) proposed an image-based disease identification method using a contrast-enhanced deep neural network [8]. The dataset included 394 Mango images in four different classes. The three diseases anthracnose, midge and powdery mildew were analyzed and the data collected was compared with a healthy mango. The authors used the contrast enhancement method to find the color contrast and the CNN method to obtain an accuracy of 88.46%.

TN Pham et. al., (2020) discussed the classification of early disease in mango leaves using hybrid metaheuristics and neural network feature selection [9]. They took 950 images in their dataset in four different classes. The three diseases anthracnose, midge and powdery mildew were tested. The data was randomly split using the Python shuffle method. The pretrained data was created using CNN models.

Anjana et al., (2019) comparatively analyzed the algorithm to identify capsicum diseases and classify whether the leaf was normal or diseased [10]. Gray Level Co-occurrence Matrix (GLCM) features were extracted for the infected area using feature inference. The diseases were classified using four different classification trees, Linear Discriminant, KNN and SVM . A total of 70 images were taken for analysis. 40 images were identified as diseased and 30 healthy images. Accuracy was 100% in SVM.

Suwit Wongsila et al., (2021) developed an algorithm for the detection of anthracnose-infected mangoes [11]. This study applied a deep learning system for classifying fungal diseases in plants. In the experimental setup, the core system was developed using CNN. In the training procedure, the dataset was divided into two parts, training and test datasets, so that 125 mango images with disease and 131 images without disease. In the efficiency test, 85 images with disease and 97 samples without disease were used. Researchers developed an ML algorithm for detecting mangoes infected with anthracnose fungal disease that yielded an accuracy of over 70%.

Chutinan Trongtorkid et al., developed an expert system for diagnosing plant diseases in mango barracuda [12]. They suggested a decision system to diagnose diseases occurring on their farm. The plant diagnostics application applied a rule-based pattern knowledge base system to a dataset of 129 leaf images. Researchers tested 16 symptom attributes of infected leaves classified into 3 class types, namely normal leaf, anthracnose, and algal spot. The accuracy of the proposed model is about 89%.

Mia M R et al., (2020) presented a Neural Network Ensemble (NNE) system for Mango Leaf Disease Recognition (MLDR). He identified the disease using the ANN, KNN and SVM

methods [13]. Features were extracted using the Gray Level Co-occurrence Matrix (GLCM) method. MLDR has been trained by 20 champions. In this case, training continued until the error rate reached 0.2.

### Objectives

After reviewing all previous literatures, this research work is part of the development of a recommendation engine to predict and pre-treat mango leaf diseases with the following objectives.

- Acquire knowledge from a real-time test bed and build an expert model using appropriate knowledge representation techniques.
- Develop a knowledge-based recommendation engine for mango disease forecast and pre-treatment.

### Methodology

The methodological way of presenting this research will give a clear view. This research focuses on the prediction and pre-treatment of mango diseases, especially anthracnose and powdery mildew, based on the recommendations provided by the rules-based reasoning approach [14,15]. Data collected through a real-time test bed on mango farms can help detect diseases and can also assess their severity. The research framework is depicted below. There are five main phases in the proposed framework.

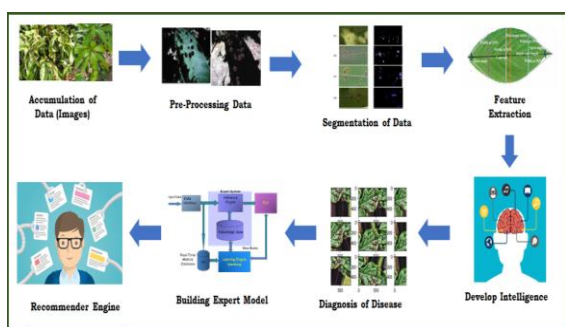


Figure 1. Framework for the recommendation engine for mango leaf disease prediction

### Phase 1: Creation of the test bed and accumulation of data

The leaf imaging process will begin with setting up a test bed for the research. Taking pictures of different types of leaves with a high-resolution camera is very necessary for good

results and efficiency. Researchers will use field IoT and smart sensor technology to collect highly relevant real-time data from mango farms [16]. The surveillance camera can be set up at the mango farms located near the research site. Three different locations can be selected within a distance of at least one kilometer with adequate security in place. A good number of healthy leaf images and diseased leaf images should be collected for research and will be continuously synced to ThingSpeak cloud storage.

### Phase 2: Pre-Processing & Segmentation of Data

Image pre-processing performs operations on images at the lowest level of abstraction, with the goal of enhancing image data by removing unnecessary distortions and enhancing critical characteristics for future processing [17]. Since the collected images have variable resolution and quality, it is essential to pre-process them in the dataset. Techniques such as contrast enhancement, scaling, scaling, angular rotation, vertical or horizontal flipping, and RGB to grayscale conversion are applied to images [18]. To change the scale or dimensions, the center cropping method is used and the images are scaled to a specific size based on the dataset. Additionally, the images are processed using a Gaussian filter, which removes noise and other unwanted characteristics from the images. The flowchart of the process involved in image pre-processing is presented in Figure 2.

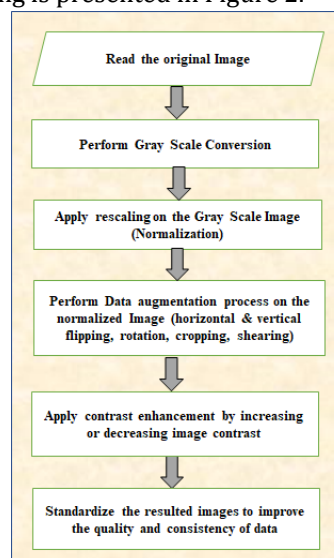


Figure 2. Flowchart of the process involved in image pre-processing

### Data segmentation

Partitioning or segmentation is usually done to get the region of interest. For the identification of the leaf disease of *Mangifera Indica*, it is done by obtaining the area in which the leaf is affected by the disease. K-means clustering groups points that are small distances from the center point of the centroid. The threshold method is another critical segmentation technique that is capable of obfuscating unnecessary data and highlighting hidden features in the dataset [19]. Data augmentation is performed on the image dataset to increase the size of the dataset by augmenting it with varying copies of the existing data present in the dataset. This process acts as a stabilizer to reduce overfitting when the training process runs. Perspective, intensity, and affine transformation are some of the commonly performed transformation techniques.

### Phase 3: Feature extraction and intelligence development

In this research, characteristics are extracted from *Mangifera Indica* leaves using the Gray-Level Cooccurrence Matrix (GLCM) method [20], which is a spatial-level dependence matrix. Co-occurrence matrix is defined as a statistical method that helps to extract second-order statistical structural features of the given image. From each cluster different shapes, textures and statistical characteristics are extracted to be used by the classifier to identify the anomalies of the given image [21].

---

#### Algorithm Feat\_Extract

---

**Input:** Leaf dataset

**Output:** Result from Neural Network model

---

Step 1: Start

Step 2: Input data

Step 3: Classify the input dataset as train and test datasets

Step 4: Reshape the data by converting it into float datatype

Step 5: Normalize the data from (0-255) to (0-1)

Step 6: Define a model function with activation function

Step 7: Execute the model and collect the error rate

Step 8: Calculate the level of accuracy of the model function

Step 9: Collect the result

Step 10: Stop

---

### Develop intelligence

Classification is an important step in classifying data into classes that help in identifying the type of disease. The five texture characteristics such as contrast, correlation, entropy, energy and homogeneity are calculated for healthy and diseased *Mangifera Indica* leaves which are used as input to the classifier [9]. SVM will be used to identify classes that are closely related to known classes [18]. It creates the optimal separating hyperplane between the classes using the training data. The extracted feature dataset is used as training data and fed as input to the SVM training classifier so that it can classify healthy/diseased *Mangifera Indica* leaf images. The various parameters, namely True Positive Rate (TPR), False Positive Rate (FPR) to verify the performance of the classifier in terms of sensitivity, specificity and accuracy, are calculated to identify infected leaves [19].

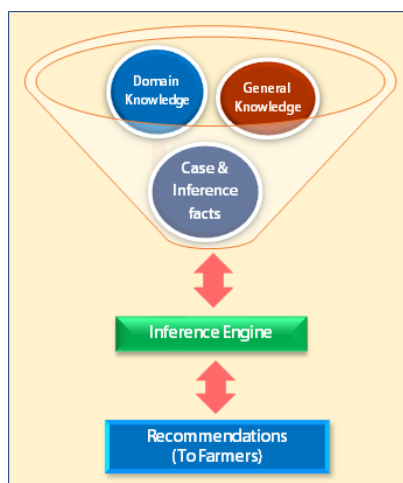
### Phase 4: Diagnosis of the disease and construction of the expert model

The K-Means clustering algorithm is used for the classification of unlabeled collected data that contains RGB values and texture characteristics. This algorithm is applied to each of the samples in a particular group denoting a deficiency class of the sampled leaf. This technique will produce an easy grouping of diseased leaves belonging to a particular class. Hybrid approaches are also used to detect an infected area on the leaf, which improves the classification process [17]. The pre-trained model is used as a base feature extractor, while the newly added layers act as a high-dimensional feature extractor responsible for classification.

### Expert Model construction

The Expert model is used to diagnose infection and the spread of disease. This model consists of a knowledge base, an inference mechanism, and a user interface. The user can answer the question in the form of 'Yes' or 'No' through the user interface, and the system will compare the user's answer with the knowledge base to identify diseases [22]. The data classification technique was developed to diagnose diseased leaves of *Mangifera Indica* using a decision tree and to diagnose a disorder

and identify its disease. An expert system based on fuzzy logic and a decision tree will produce high accuracy.



Phase 5: Develop the recommendation engine

A recommendation engine uses specialized algorithms and machine learning solutions to offer recommendations to users. Powered by the automated setup, coordination, and management of machine learning's predictive analytics algorithms, the recommendation engine can wisely select which filters to apply to a particular user's specific situation. Regression techniques such as linear regression, random forest regression, polynomial regression, decision tree regression and support vector regression will be applied to offer a solution to the mango farm farmers in the specific critical situation [24]. The prediction accuracy must be kept high to predict infection on mango leaves and the spread of any new disease. Finally, the majority voting (MV) method will be applied as a combination technique to provide the final accuracy [25].

### Significance Of the Research

The recommendation engine developed from this research alleviates the problem of limited number of experts in the preliminary diagnosis and treatment of mango diseases and motivates further research in the agricultural expert system area. The immediate beneficiaries of the research are major growers of *Mangifera Indica* and agricultural professionals or agronomists. In particular, the expert system will be of great importance in teaching primary agricultural educators, general agronomy experts to

understand mango diseases well. As a result, mango growers can use the system to diagnose mango diseases in primary agricultural sectors where highly trained professionals are not available. Additionally, the recommendation engine can be used as a guide for agricultural professionals.

### Social Relevance Of The Research

According to the report of the Agricultural and Processed Food Export Development Authority (APEDA) dated March 2022, the demand for mango products is high in India [26]. During the year 2019-2020, the country exported 85,725.55 tonnes of mango pulp to the world, worth 584 crores/US\$81.88 million. But diseases cause 10-70% fruit loss and 10-85% yield loss, which has had a big impact on production and annual income. As Theni district is one of the largest contributors to mango cultivation in Tamil Nadu, this research will help farmers take on the task of predicting infection, determining mango yield and increasing production. It is most effective in helping farmers make a decision and analyze an agricultural problem when an expert is not available. With cameras placed in the farm, farmers can monitor their mango trees from home. They can easily follow the progress of their farm from a computer or even a mobile device.

### Conclusion

Horticulture is one of the many industries that has been altered by machine learning. The Mango Growers of Western Ghats of India can optimise disease growth, use fewer resources, and increase the sustainability of their operations with the aid of machine learning technologies. In the absence of highly qualified professionals, this research study proposes a system for diagnosing mango illnesses in primary agricultural areas. Also, the recommendation engine can be utilised as a guide for agricultural specialists. It will be useful for infection forecasting, figuring out mango yield, and raising output. Moreover, it functions as an aid to the farmers in decision making and assess an agricultural problem when an expert is not available.

### Acknowledgement

The authors would like to thank the management of Jayaraj Annapackiam College for Women (Research grant No. JAC/ 2022-2023/JACFRP – Faculty/ 8) for the financial support of this research.

### References

1. The Wall Street Journal, June 18, 2022 <https://miramw.org/market-trends/heat-wave-devastates-indias-mango-harvest/>
2. Samiksha S, "Mango Cultivation in India – Production Area, Climate, Harvesting and Fruit Handling", <https://www.yourarticlelibrary.com/fruits/mango-cultivation-in-india-production-area-climate-harvesting-and-fruit-handling/24699>, Retrieved on 26.07.2022.
3. District G'-Return -Fasli 1430, Theni, District Statistical Hand BOOK 2020-21, page 34.
4. Swetha, K., Venkataraman, V., Sadhana, G.D. and Priyatharshini, R., "Hybrid approach for anthracnose detection using intensity and size features", IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR), 2016, pp. 28-32. IEEE.
5. S. Arya and R. Singh, "A Comparative Study of CNN and AlexNet for Detection of Disease in Potato and Mango leaf," International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT), GHAZIABAD, India, 2019, pp. 1-6, doi: 10.1109/ICICT46931.2019.8977648.
6. Mustafa Merchant, Vishwajeet Paradkar, Meghna Khanna and Soham Gokhale,"Mango Leaf Deficiency Detection Using Digital Image Processing and Machine Learning", 3rd International Conference for Convergence in Technology (I2CT), The Gateway Hotel, Pune, India., 2018
7. Prakash, B. and Amit Yerpude. "Identification of Mango Leaf Disease and Control Prediction using Image Processing and Neural Network." 2015.
8. Kien Trang, Long TonThat, Nguyen Gia Minh Thao and Ngoc Tran Ta Thi, "Mango Diseases Identification by a Deep Residual Network with Contrast Enhancement and Transfer Learning", 978-1-7281-3276-1/19/\$31.00 ©2019 IEEE, pp. 138 – 142.
9. T. N. Pham, L. V. Tran and S. V. T. Dao, "Early Disease Classification of Mango Leaves Using Feed-Forward Neural Network and Hybrid Metaheuristic Feature Selection," IEEE Access, Volume 8, pp. 189960-189973, 2020, doi: 10.1109/ACCESS.2020.3031914.
10. Anjna, Meenakshi Sood, Pradeep Kumar Singh,"Hybrid System for Detection and Classification of Plant Disease Using Qualitative Texture Features Analysis", Procedia Computer Science, Volume 167, 2020, pp. 1056–1065
11. Suwit Wongsila, Parinya Chantrasri and Pradom Sureephong, "Machine Learning Algorithm Development for detection of Mango infected by Anthracnose Disease", Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunication Engineering, 2021 IEEE, DOI: 10.1109/ECTIDAMTNC51128.2021.9425737
12. Chutinan Trongtorkid and Part Pramokchon, "Expert System for Diagnosis Mango Diseases Using Leaf Symptoms Analysis", The 3rd International Conference on Digital Arts, Media and Technology (ICDAMT2018), pp. 59 –64, DOI: 978-1-5386-0572-1/18/\$31.00 ©2018 IEEE
13. Md. Rasel Mia, Sujit Roy, Subrata Kumar Das, Md. Atikur Rahman, "Mango leaf disease recognition using neural network and support vector machine", Iran Journal of Computer Science, 2020, pp. 1-12, <https://doi.org/10.1007/s42044-020-00057-z>
14. R. A. JM. Gining, S. S. M. Fauzi, N. M. Yusoff, T. R. Razak, M. H. Ismail, N. A. Zaki and F. Abdullah, "Harumanis mango leaf disease recognition system using image

- processing technique", Indonesian Journal of Electrical Engineering and Computer Science, Volume 23, Issue 1, 2021, pp. 378~386, ISSN: 2502-4752, DOI: 10.11591/ijeecs.v23.i1.pp378-386.
15. R Shantha Mary Joshitta and L Arockiam, "A Predictive Model to Forecast and Pre-Treat Diabetes Mellitus using Clinical Big Data in Cloud", International Journal of Applied Engineering Research", International Journal of Applied Engineering Research, Volume 10, Issue 82, 2015, pp. 55-59
  16. A Jenifer Jothi Mary, S. Santiago Sebastian, Dr. L. Arockiam, "A Methodological Framework to Identify the Students' Opinion using Aspect based Sentiment Analysis", International Journal of Engineering and Technical Research, Volume 5, Issue 02, February 2016, DOI: 10.17577/IJERTV5IS020528
  17. Balasundaram A, Pradeep KV, Sandhya S, "An Extensive Study on Disease Prediction in Mango Trees Using Computer Vision", Annals of R.S.C.B., ISSN:1583-6258, Volume 25, Issue 1, 2021, Pages. 1895 - 1905
  18. Trang, K., TonThat, L., Thao, N.G.M. and Thi, N.T.T., "Mango Diseases Identification by a Deep Residual Network with Contrast Enhancement and Transfer Learning", In the proc. of IEEE Conference on Sustainable Utilization and Development in Engineering and Technologies (CSUDET), 2019, pp. 138-142.
  19. P. G. Palanimani, J. J. M. A, V. S. Kumar, S. B. Kasturiwala, S. Ahmareen and S. Bohra, "A novel approach for detection of diabetes patients using machine learning techniques," International Interdisciplinary Humanitarian Conference for Sustainability (IIHC), Bengaluru, India, 2022, pp. 1402-1408, doi: 10.1109/IIHC55949.2022.10059675.
  20. G. M. Karthik, S. Mary Joshitta, R. R, B. Manjulatha, J. -k. Chung and L. P. Gagnani, "Design of Covid19 Detection Based on Relative Eccentric Feature Selection Using Deep Vektored Regressive Neural Network for Corona Virus," International Interdisciplinary Humanitarian Conference for Sustainability (IIHC), Bengaluru, India, 2022, pp. 1027-1033, doi: 10.1109/IIHC55949.2022.10060368.
  21. Sagra Vetal and R.S. Khule, "Tomato Plant Disease Detection using Image Processing." International Journal of Advanced Research in Computer and Communication Engineering, Volume 6, Issue 6, 2017, pp. 293-297.
  22. Manjunath Hosmani, Nithesh. J, Akbarbasha H Nadaf and Jeeshan In, "Detection of unhealthy region of plant leaves and Classification of plant leaf diseases using texture features." International Journal of Emerging research in management & technology, Volume 6, Issue 5, 2017, pp. 156-162.
  23. K. Renugambal B. Senthilraja, "Application of Image processing techniques in plant disease recognition." International Journal of Engineering Research & Technology, Volume 4, Issue 3, 2015, pp. 919-923.
  24. Mamata Garanayak, Sachi Nandan Mohanty, Alok Kumar Jagadev and Sipra Sahoo, "Agricultural Recommendation System for Crops Using Different Machine Learning Regression Methods", International Journal of Agricultural and Environmental Information Systems, Volume 12, Issue 1, 2021, pp.93-101,
  25. Mamata Garanayak, Sipra Sahoo, Sachi Nandan Mohanty and Alok Kumar Jagadev, "An Automated Recommender System for Educational Institute in India", EAI Endorsed Transactions on Scalable Information Systems, Volume 7, Issue 26, e9, 2020, pp. 1-13. Doi: 10.4108/eai.13-7-2018.163155
  26. Global Processed Mango Products Market Size, Share & Trends Analysis Report by Type (Dried/Dehydrated Mangoes, Puree, Pulp & Concentrate, Fruit Bars, Candies, Juice, Pickles, and Others), and by Distribution Channel (Online and Offline) Forecast Period (2022-2028), Published: Apr 2022, Report Code: OMR2026416